Journal of Memory and Language 72 (2014) 49–58

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Statistically coherent labels facilitate categorization in 8-month-olds

Lucy C. Erickson [a,*], Erik D. Thiessen [a], Katharine Graf Estes [b]

[a] Carnegie Mellon University, United States
[b] University of California, Davis, United States

ARTICLE INFO

Article history:
Received 16 August 2013
revision received 22 November 2013
Available online 29 January 2014

Keywords:
Statistical learning
Word segmentation
Labels
Categorization

ABSTRACT

There is considerable evidence that infants can segment speech using syllable co-occurrence probabilities; however, relatively less is known about the nature of the representations formed during this process. The present studies tested the prediction that statistically segmented items should exhibit a specific property of real words, namely, these items should have a facilitative effect on infant categorization. During the segmentation phase, eight-month-old infants listened to a fluent speech stream that contained statistical word boundary cues. Infants were then tested on their ability to categorize drawings of an unfamiliar category when category exemplars were paired with either high-probability or low-probability labels from the segmentation phase. Infants who heard high-probability labels showed evidence of categorization. In contrast, infants who heard low-probability labels did not. A follow up experiment revealed that this effect was due to facilitation for high-probability words rather than inhibition for low-probability items. These results fit with theoretical accounts that suggest that infants treat statistically segmented units as potential words.

Introduction

A critical task that infants face in the first years of life is to acquire a lexicon. This task is challenging because infants hear relatively few words in isolation (Brent & Siskind, 2001), and because word boundaries in fluent speech are marked by a combination of imperfect cues rather than a single infallible cue (Cole & Jakimik, 1980). Consequently, the question of how infants accomplish word segmentation has generated a large body of research (e.g., Jusczyk & Aslin, 1995; Jusczyk, Houston, & Newsome, 1999; Saffran, Aslin, & Newport, 1996; for a review, see Jusczyk, 1999). One possibility is that infants rely on a cross-linguistically consistent characteristic of words, namely, that syllables that comprise words cluster together with greater reliability than syllables that occur incidentally across word boundaries (Harris, 1955; Hayes & Clark, 1970). Thus, a mechanism that infants may use to segment their first words from continuous speech is statistical word segmentation, the process of segmenting units from speech on the basis of syllable co-occurrence probabilities.

Statistical learning is a powerful learning mechanism that is available from birth to adulthood (Graf Estes, Evans, Alibali, & Saffran, 2007; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). In statistical word segmentation experiments, learners require only a brief familiarization with a fluent speech passage to track syllable patterns that identify words (high probability clusters) versus across-word sequences (low probability clusters). Moreover, because this mechanism does not depend on language-specific knowledge, it may be of particular use in infants' initial discovery of word forms. As such, it may be the case that statistical learning plays an important role in early lexical acquisition (e.g.,

* Corresponding author. Fax: +1 412 286 2798.
E-mail address: lerickso@andrew.cmu.edu (L.C. Erickson).

0749-596X/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2014.01.002

Saffran & Thiessen, 2007; Thiessen & Saffran, 2007; Werker & Curtin, 2005). Consistent with this possibility, statistical learning appears to be related to various aspects of lexical development (e.g., Evans, Saffran, & Robe-Torres, 2009; Romberg & Saffran, 2010).

However, there have been objections to the claim that statistical learning contributes to language development. One centers on the fact that most studies of statistical learning have been conducted in laboratory settings with artificial stimuli. This has raised questions about whether the learning processes that are found in the laboratory are the same as those involved in developing a lexicon in infants' natural environments (e.g., Graf Estes, 2012; Johnson & Tyler, 2010). A related objection is that the representations that emerge from statistical learning are not linguistic or lexical in nature. For example, Endress and Mehler (2009) assert that statistical word segmentation does not result in word-like representations. Rather, they have proposed that although learners are able to compute transitional probabilities between syllables, there is no evidence that "the items with stronger [transitional probabilities] are represented as actual word-like units, or even that they have been extracted" (p. 352). Instead, they have argued that statistical learning produces chaining memories that correspond to syllable transitions, but does not result in the storage of an integrated percept (but see Perruchet & Poulin-Charronnat, 2012). These issues are problematic for accounts that propose that infants use statistical segmentation to discover their first word forms. If infants do use statistical segmentation for lexical acquisition, they should treat the output of statistical learning as potential word forms. The goal of the current research was to test whether the product of statistical learning exhibits an established property of real words, which exert an influence on infant object categorization.

To address questions about the nature of the representations emerging from statistical learning, researchers have attempted to determine whether they exhibit the properties of real words. One approach to this question has been to assess whether computational models of statistical learning give rise to representations that are similar to those thought to exist in the lexicon. For example, chunking models of statistical learning give rise to the kinds of unitary representations that are thought to characterize lexical items (e.g., Perruchet & Vinter, 1998; Giroux & Rey, 2009; Orbán, Fiser, Aslin, & Lengyel, 2008; Thiessen, Kronstein, & Hufnagle, 2013). A complementary approach to this question is to use behavioral methods to probe the representations resulting from infant statistical learning. In one such study, Saffran (2001) tested whether preferences for statistical nonsense words parsed from fluent speech (e.g., *tibudo*) differed as a function of the lexical context. These nonsense words were either embedded into a highly familiar English sentence frame (e.g., "I like my tibudo") or one that was a nonsense sentence matched on several dimensions (e.g., "Zy fike ny tibudo"). Eight-month-old infants listened longer to sentences that included statistically coherent words from the speech stream than to sentences containing sequences that crossed word boundaries. However, this preference only held when the words were presented in the context of a real English sentence. There was no difference in listening times for infants who were exposed to nonsense word frames, which indicates that they only treated these items differently in meaningful linguistic contexts.

Graf Estes et al. (2007) used a word learning paradigm to investigate whether the process of statistically segmenting words from fluent speech is related to the process of mapping meanings to labels. The authors found that 17-month-old infants were able to map labels to objects when those labels comprised syllable sequences with high internal probabilities, based on prior statistical segmentation experience. In contrast, infants did not learn the mapping when the labels comprised familiar sequences with low internal probabilities. This indicates that the process of segmenting words from fluent speech is intimately linked to word learning. Because infants had already segmented these labels and stored them in memory, cognitive resources that might otherwise have been devoted to learning the word form were free to learn the mapping between the label and its referent. Taken together, Saffran (2001) and Graf Estes et al.'s (2007) studies suggest that infants treat high probability items, and thus the output of statistical learning, as possible native language words rather than merely sound sequences with high internal probabilities but no linguistic relevance.

We hypothesized that if the output of statistical learning is indeed word-like, it should exhibit additional properties of real words. One such property relates to the relationship between linguistic labels and infant categorization. A body of research has explored the effects of linguistic and nonlinguistic labels on infant object categorization (e.g., Balaban & Waxman, 1997; Ferry, Hespos, & Waxman, 2010; Fulkerson & Haaf, 2003; Namy & Waxman, 1998). In this literature, infants are familiarized with exemplars of novel object categories (e.g., rabbits) paired with a consistent label (e.g., "A *toma*" or "Look at the *toma*"). They are subsequently tested on their knowledge of the category as a whole as indexed by their ability to discriminate a novel exemplar of the familiar category (i.e., a rabbit) from a novel exemplar of a new category (e.g., a pig). Research using this procedure has found that the types of symbols that are conducive to category formation tend to be symbols that are used to refer to objects in infants' real environments (Balaban & Waxman, 1997; Ferry et al., 2010; Fulkerson & Waxman, 2007; Namy & Waxman, 1998; Woodward & Hoyne, 1999). For example, infants readily form an object category (e.g., dinosaurs) when exemplars of that category are consistently accompanied by a linguistic label; when the exemplars are accompanied by tones they do not exhibit categorization (e.g., Balaban & Waxman, 1997; Ferry et al., 2010; Fulkerson & Waxman, 2007).

The relationship between label identity and categorization performance is not static over development. Instead, the range of signals that permit categorization appears to begin relatively broadly and to narrow over the course of development, as infants become attuned to their native language. A recent study demonstrated that 3- to 4-month-old infants categorize in the presence of nonhuman primate vocalizations (Ferry, Hespos, & Waxman, 2013), similar to their performance with linguistic naming

phrases (Ferry et al., 2010). By 6-months, however, nonhuman primate vocalizations no longer seemed to be conducive to category formation (Ferry et al., 2013).

By our perspective, because auditory and visual information compete for attentional resources, a critical factor that determines whether infants will categorize is the familiarity of the label or auditory associate (e.g., Robinson & Sloutsky, 2007). Early in development, infants possess a broad category of what constitutes a familiar utterance. For very young infants, this category may even include nonhuman primate vocalizations (Ferry et al., 2013). As infants accrue experience with their native languages, however, this category is refined and becomes more narrow. At this point, primate vocalizations are no longer contained within this category and are no longer treated as familiar. Instead, they are likely to interfere with categorization by drawing attention away from visual processing (e.g., Robinson & Sloutsky, 2007).

Thus, the familiarity of auditory associates influences visual categorization (e.g., Robinson & Sloutsky, 2007). Because research suggests that statistical learning results in familiar representations (Graf Estes et al., 2007), a prediction that follows is that statistical learning will produce representations that facilitate categorization. In the first of two experiments, we tested whether the facilitation in object categorization occurs for syllable sequences that infants have segmented from fluent speech using statistical cues, which should now be familiar to the infants if they have been stored in memory as potential words. We examined categorization outcomes as a function of the statistical properties of the accompanying labels. For some infants, the category exemplars were accompanied by a statistically-defined label with high internal probability; for other infants the label was a low probability sequence that crossed word boundaries in the language, termed a partword. If statistical word segmentation produces word-like units, high transitional probability sequences should be familiar to the infants and thus facilitate categorization, just as other lexical items do. In contrast, partwords have not been culled from speech and are thus unfamiliar the infants. In the second experiment, we addressed additional methodological concerns. The logic of these experiments was built upon the domain-general premise that category learning should be improved if exemplars are paired with a well-encoded percept (e.g., a word). Here, statistical learning should provide infants with a well-encoded, familiar perceptual unit to facilitate categorization.

We chose to investigate this question with 8-month-old infants for several key reasons. First, these infants are old enough to demonstrate behavioral segmentation based on statistical structure (e.g., Saffran et al., 1996). In addition, they know the meanings of very few words (e.g., Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999), but they have just begun to segment words from speech and map them to referents. This is critical, because if statistical learning really yields representations that are linguistic in nature, these representations should be linguistic from the beginning. That is to say, the output of statistical learning should be word-like before lexical development has created expectations about what it means to be a word, and before the fully developed lexicon has been established. If statistical segmentation plays a fundamental role in lexical development, its influence must be in place as the lexicon is forming. The outcome of these studies has important implications for understanding the nature of the representations formed during statistical word segmentation and their influence on learning label-category mappings. If we can establish that these processes are in place early in development, when infants are on the cusp of building a lexicon, this will provide further support for the role of statistical segmentation in language acquisition.

Experiment 1

In this experiment, we familiarized 8-month-old infants with a nonsense language produced in fluent sentences by a human speaker. The statistical structure of the language was such that some syllable clusters appeared together with 100% reliability whereas others appeared together with only 25% reliability. The high probability clusters formed words in the language; the low probability clusters occurred across word boundaries. Following familiarization, infants underwent a categorization task that used the same visual stimuli and design as previous studies (line drawings of dinosaurs and fish; Balaban & Waxman, 1997; Ferry et al., 2010). The critical manipulation was whether labels that accompanied the exemplars were words or partwords during the segmentation phase.

We predicted that if the output of statistical word segmentation is truly word-like, infants should successfully form a visual object category in the presence of syllable clusters that have high internal probabilities during segmentation. In contrast, we predicted that syllable clusters that had low internal probabilities during segmentation—which should not be treated as candidate word forms—would not lead to successful categorization. However, it is also possible that statistical learning does not produce potential word forms and instead results in syllable clusters with high internal probabilities that have no lexical status (Endress & Mehler, 2009). If this is the case, there should be no difference in infants' categorization in the presence of low- and high-probability syllable clusters. Based on prior research, we predicted 8-month-olds would express successful categorization as a novelty preference for the exemplar from the never-viewed category (e.g., Ferry et al., 2010).

Method

Participants

We tested 36 8-month-old infants (18 females). All infants were born full term, and they had no prior history of vision or hearing impairments. Eighteen infants (9 females) were randomly assigned to the word condition (labels comprised high probability syllable sequences; $M = 8.16$ months, Range = 7.80–8.40) and 18 infants (9 females) were randomly assigned to the partword condition (labels comprised low probability syllable sequences; $M = 8.16$ months, Range = 7.76–8.50). Sixteen additional infants were tested and replaced due to fussiness

($n = 12$), playing with a toy during the experiment ($n = 1$), computer failure ($n = 1$), or moving out of the camera frame ($n = 2$). Participants were recruited from a small west coast city and received either a book or a t-shirt in appreciation for their time.

Stimuli

Auditory stimuli. The auditory stimuli used in the segmentation and category familiarization phases were originally designed by Thiessen, Hill, and Saffran (2005). The segmentation phase consisted of a nonsense language produced in fluent, infant-directed sentences by a native English speaker who was naïve to the statistical structure of the language. The 1.02 min language included 12 acoustically distinct sentences, each of which each started with the syllable "mo" and ended with the syllable "fa" so that the silences between sentences could not be used as a cue to word boundaries. Each sentence contained a pseudo-randomized order of four nonsense words (*dibo, kuda, lagoti, nifopa*). They were spoken at an average rate of 2.5 syllables per second and 1.3 sec of silence separated the sentences. The average pitch (measured by fundamental frequency, F0) of the speaker was 292 Hz, with a range of 140–480 Hz. This large range reflects the exaggerated pitch contours of infant-directed speech. These values are consistent with previous work on the characteristics of infant-directed speech (e.g., Fernald, 1989).

Previously, Thiessen et al. (2005) ruled out the possibility that these sentences contained acoustic cues to word boundaries by collecting judgments from adults. The participants listened to a random sentence from the language as many times as desired, then indicated which of two items (a word and a foil) was more likely to have been a word in that sentence. Performance was at chance; the participants did not reliably detect the words from listening to individual sentences. The authors concluded that cues in addition to sequential syllable orders did not inform word boundaries.

The same speaker recorded the two test items, the word label *lagoti*, and the partword label, *danifo*, which constituted the end of the word *kuda* and the beginning of the word *nifopa*. The labels were produced in isolation.

Visual stimuli. The visual stimuli for the categorization task were identical to those used by Fulkerson and Waxman (2007). A pool of 20 slides of colored line-drawn dinosaurs and fish were used to create familiarization sets of eight exemplars within each category, and one additional category exemplar was paired with a novel category exemplar for the test trial set. To control for potential item effects, stimuli were randomized such that individual exemplars occurred in both familiarization and test sets for different infants. During test trials, right and left position of the exemplars was counterbalanced across infants. Images on the screen were displayed at 12.25 × 9.5 in.

Procedure

Infants were tested individually in a sound-attenuated testing room, seated on a caregiver's lap approximately 42 in. away from a 35.3 × 20.5 in. LCD monitor with integrated speakers. An experimenter outside the testing room observed the infant over closed-circuit video and controlled the presentation of the stimuli using Habit X software (Cohen, Atkinson, & Chaput, 2004). To eliminate bias, parents wore headphones playing music, and the experimenter was blind to the nature of the stimuli being presented. The procedure included a segmentation phase, a categorization phase, and test phase.

Segmentation phase. Infants listened to the nonsense language while viewing a series of 6 still images from a nature-themed screen saver (e.g., close-ups of flowers). The images changed after an average of 14 s, and were used to help maintain infants' attention during segmentation.

Category familiarization phase. Immediately following the segmentation phase, infants participated in a novelty-preference task to measure categorization. The design was based on the task used previously by Balaban and Waxman (1997) and Ferry et al. (2010). Infants viewed a series of eight different exemplars of either dinosaurs or fish (counterbalanced across participants) presented for 10 s each. Each exemplar was a distinct color, and was accompanied by a consistent linguistic label (i.e., *lagoti* or *danifo*). For the infants in the word condition, the syllables that comprised this label (*lagoti*) co-occurred with 100% reliability during the segmentation phase. For infants in the partword condition, the syllables in the label (*danifo*) only occurred together 25% of the time, because they co-occurred incidentally between word boundaries. Each label played 3 s after the trial onset, and repeated after a 3 s pause (see Fig. 1 for a sample schematic). Between trials, infants viewed an animated cartoon clip in order to direct their attention to the screen presenting the exemplars.

Categorization test phase. During this phase, infants viewed a silent test trial, presented for 10 s. Regardless of whether infants were familiarized with fish or dinosaurs, each test trial consisted of a dinosaur and a fish presented simultaneously, neither of which had been seen during categorization. During test trials, the fish and dinosaur pair were the same color. Previously, infants have failed to discriminate these two categories at test following familiarization to the exemplars in the presence of tones, but they succeeded (i.e., showed evidence of discrimination of the two category exemplars) following familiarization to the exemplars in the presence of linguistic labels (e.g., Balaban & Waxman, 1997; Ferry et al., 2010). We tested whether statistically-defined word and partword labels (both instances of human speech) facilitate categorization equally, or whether certain kinds of human utterances confer advantages in categorization over others.

Coding

Trained observers, blind to the hypotheses and nature of the stimuli, coded infant visual fixations offline to determine left, right, and central looking using specialized software (see Fernald, Zangl, Portillo, & Marchman, 2008, for further information). One third of infants in each condition were recoded by an independent observer to obtain a measure of reliability. The coders agreed (within 1 frame) on 99.7% of the overall frames and 97.5% of the frames in

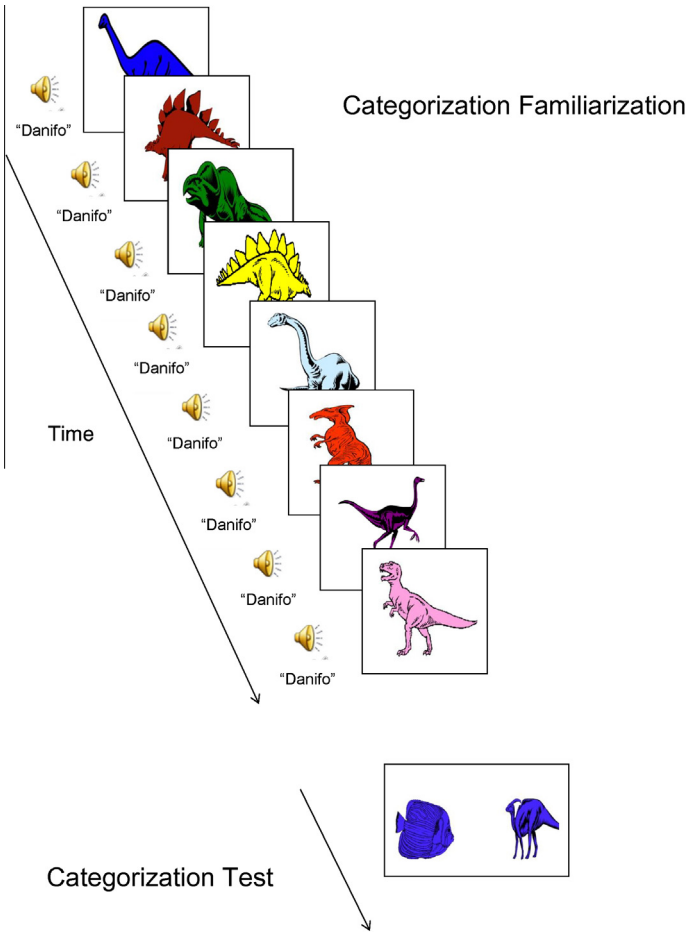

Fig. 1. Experimental paradigm: Sample schematic of categorization familiarization and test paradigm.

which a shift in looking location occurred. As described in more detail below, we used this coding to analyze infants' looking during the familiarization and test trials.

Results and discussion

Category familiarization

We first examined the mean proportion of total trial time during familiarization that infants spent attending to the exemplars in each condition. An independent samples *t*-test revealed that infants in the word condition ($M = 0.83$, $SD = 0.14$) looked significantly longer than infants in the partword condition ($M = 0.73$, $SD = 0.14$), $t(34) = 2.30$, $p = .03$, Cohen's $d = 0.74$. Infants who heard words presented with category exemplars tended to spend more time fixating the exemplars during familiarization than infants who heard partwords.

Categorization test

Consistent with previous studies, a novelty preference score was used as the index of categorization. It was calculated by dividing the total duration of looking time to the novel category exemplar (e.g., the fish when the infant was familiarized with dinosaurs) by the duration of total looking time to both exemplars (e.g., looks to the fish plus looks to the dinosaur). Based on prior experiments, we predicted that if infants categorized successfully, they would show a novelty preference for the novel category exemplar.

Novelty preference scores for the two groups were compared using an independent samples *t*-test. This analysis revealed that infants in the word condition ($M = 0.59$, $SD = 0.15$) displayed larger novelty preference scores than infants in the partword condition ($M = 0.46$, $SD = 0.12$), $t(34) = 2.84$, $p = .008$. The effect size for this analysis ($d = .96$) was found to exceed Cohen's (1988) convention for a large effect ($d = .80$).

We also investigated whether infants in each group showed looking patterns that were reliably different from chance level (.5, or 50% looking to each item). As illustrated in Fig. 2, infants in the word condition displayed a significant novelty preference for the exemplar from the category not viewed during familiarization, one sample $t(17) = 2.64$, $p = .017$, Cohen's $d = 0.62$ (13 of 18 infants showed a novelty preference). Infants in the partword condition did not exhibit looking patterns that differed statistically from chance, and in fact showed a slight preference for the familiar, rather than novel, test item, $t(17) = 1.28$, $p = .22$, Cohen's $d = 0.30$ (only 9 of 18 infants showed a novelty

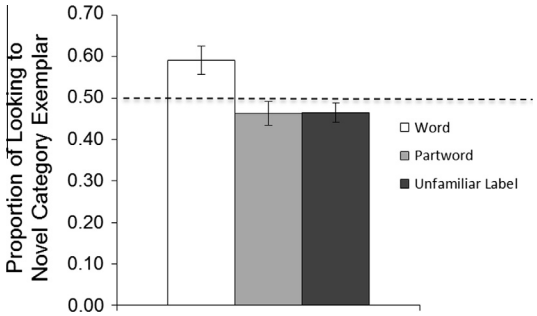

Fig. 2. Infants' looking times to familiar and novel category exemplars during Experiment 1 and 2 by label type. Dashed line indicates chance performance. Values higher than dashed line represent a novelty preference whereas values below the dashed line represent a familiarity preference. Error bars represent S.E.M.

preference). This indicates that infants who heard statistically-defined words during familiarization with the exemplars of a category discriminated the two categories at test. In contrast, the infants who heard low-probability partwords during familiarization did not discriminate the two categories.

Experiment 2

In Experiment 1, we found that the labels that infants detected in a statistical word segmentation task promoted categorization more effectively than labels that crossed word boundaries during segmentation. Experiment 2 investigated the underlying cause of this difference in performance between words and partwords. One possibility, consistent with our predictions, is that experience with the segmentation phase allowed infants to extract high probability units that became strong candidate labels, and these units facilitated categorization performance (e.g., Graf Estes et al., 2007). By this account, high probability units facilitate categorization, not merely by virtue of being linguistic, but because they are familiar to the infant. An alternative possibility is that infants' experience during the segmentation phase led to inhibition of learning for the partwords, because infants have experience hearing these sequences across word boundaries. Thus, these items have low statistical coherence and make poor labels (see Mirman, Magnuson, Graf Estes, & Dixon, 2008, for similar findings with adults). Thus, it may not be that performance with statistical words was elevated, but rather that performance with statistical partwords was actually hindered. The results of Experiment 1 are ambiguous, in that they do not indicate whether superior categorization with words was due to facilitation for words, or inhibition for partwords.

To differentiate between facilitation and inhibition, in Experiment 2 we assessed infant categorization when the labels were unfamiliar: linguistic labels that did not previously occur during a segmentation phase. We hypothesized that if statistical words facilitate category formation, then infants who hear unfamiliar labels should perform similarly to infants who heard partwords in Experiment 1 because neither of these labels have prior statistical segmentation support. Such a result would suggest that linguistic status alone is insufficient to promote categorization. In contrast, if linguistic status alone promotes categorization, except in cases when a linguistic item is a poor label, infants who hear unfamiliar labels should categorize successfully, like infants who heard word labels in Experiment 1. This would indicate that the categorization failure of infants who heard partwords in Experiment 1 occurred because the statistical structure of partwords inhibited category formation. Thus, we tested categorization performance with unfamiliar labels in Experiment 2. Because unfamiliar labels are novel syllable sequences, they have no prior history to produce either inhibition or facilitation. To control for factors such as fatigue and hearing a statistically structured auditory stream, infants in Experiment 2 received familiarization with a statistically structured sequence of pure tones that had the same duration as the language used in Experiment 1.

In addition, Experiment 2 served a secondary purpose. Because the identity of the particular test items was not counterbalanced (i.e., *lagoti* was always the word and *danifo* was always the partword), the current experiment provided an opportunity to evaluate whether *lagoti* was a more compelling label to infants than *danifo* (e.g., if one of the two labels were superior in terms of English phonotactic probabilities). This alternative explanation for the results of Experiment 1 is unlikely because prior experiments using this language have found no evidence of item-based preferences (Thiessen et al., 2005). Regardless, the current experiment provided a conservative test of this alternative explanation.[1] If infants in Experiment 1 exhibited an item specific preference, that preference should persist in Experiment 2. However, if the preference for *lagoti* was contingent upon its role in the segmentation phase, no differences between the two labels should be observed when infants hear tones during the segmentation phase.

Method

Participants

We tested 18 infants (8 females; *M* age = 8.10 months, Range = 7.70–8.43) using the same inclusion criteria as used in Experiment 1. Fourteen additional participants were tested and replaced due to fussiness ($n = 11$), experimenter error ($n = 1$), parental interference ($n = 1$), and a corrupted video recording ($n = 1$).

Procedure

The procedure and stimuli were identical to Experiment 1, with the exception of the segmentation phase. To equate

[1] Although Experiment 2 provides a control for item-based effects, it does not control for segmentation-based effects, that is, effects that may arise if some items are inherently easier to cull from fluent speech than other items due to acoustic factors such as coarticulation. To control for such effects it is necessary to counterbalance the lexical status of particular items across two languages, which was not done here. However, given that the acoustic pre-testing in Thiessen et al. (2005) found no differences in the ease with which adults identified test items in the input, there is no reason to expect these kinds of segmentation-based effects play a role in the learning of this particular stimulus set.

for factors related to exposure to a statistically structured stream, the segmentation phase of Experiment 2 was designed so that infants still had the experience of segmenting units from a structured stream, but units that were not related to the lexical items they heard during the categorization phase. Infants listened to a structured sequence of pure tones (Saffran, Johnson, Aslin, & Newport, 1999). There were four "words," each consisting of a triplet of tones. Transitional probability patterns indicated boundaries between tone words. The tone sequence was matched to the language used in Experiment 1 for the duration of sentences, the duration of pauses between sentences, and overall duration. The segmentation phase was designed to be similarly taxing to the segmentation phase of Experiment 1, without providing any information that could facilitate or inhibit learning of the categorization labels. Subsequently, infants experienced the same lexical items during the category familiarization phase (either *lagoti* or *danifo)*. However, because the familiarization stimuli were tones, these labels were now completely unfamiliar.

Results and discussion

Category familiarization

To determine whether infant looking at the exemplars during familiarization differed across experiments, we performed two independent samples *t*-tests. The average looking proportion during familiarization for infants in the unfamiliar label condition ($M = 0.80$, $SD = 0.11$) did not differ from infants in the word condition of Experiment 1, $t(34) = 0.69$, $p = .49$, Cohen's $d = 0.23$, nor from infants in the partword condition, $t(34) = 1.88$, $p = .07$, Cohen's $d = 0.63$.

Categorization test

As shown in Fig. 2, a one sample *t*-test revealed that the novelty preference scores of infants who heard unfamiliar labels ($M = 0.46$, $SD = 0.10$) did not differ statistically from chance levels, $t(17) = 1.53$, $p = .15$, Cohen's $d = 0.36$. (12 of 18 showed a familiarity preference). Independent samples *t*-tests showed that the novelty preference scores of infants who heard unfamiliar labels differed significantly from infants who heard word labels in Experiment 1, $t(34) = 3.04$, $p = .004$, Cohen's $d = 1.02$. In contrast, the performance of infants who heard unfamiliar labels did not differ from infants who heard partwords in Experiment 1, $t(34) = .05$, $p = .96$, Cohen's $d = .03$.

Infants in the word and unfamiliar label conditions showed similar patterns of attention during familiarization. However, during testing, the categorization of infants who heard unfamiliar labels (Experiment 2) was similar to infants who heard partword labels rather than word labels (Experiment 1). These results indicate that the superior performance of infants who heard words instead of partwords in Experiment 1 was due to facilitation and not to inhibition. Thus, the presence of familiar and statistically coherent labels leads to categorization success above and beyond both novel labels and statistically incoherent items.

In addition, infants who heard the label *lagoti* ($M = 0.46$, $SD = .09$) showed equivalent novelty preferences to infants who heard *danifo* ($M = 0.46$, $SD = 0.11$), $t(16) = .053$, $p = .96$, Cohen's $d = .03$. Differences in categorization performance in Experiment 1 could not be attributed to a pre-existing preference for one of the two labels.

General discussion

The present research was designed to determine whether statistical word segmentation produces word-like representations. One established feature of real words is that their presence facilitates infant object categorization (e.g., Ferry et al., 2010). We found that after statistical segmentation of a nonsense language, infants categorized in the presence of labels that were statistically coherent syllable clusters from the language. In contrast, infants failed to categorize when they heard labels that comprised syllable clusters that had occurred only incidentally across word boundaries. A follow up study tested categorization in the presence of completely novel labels. It demonstrated that the superior performance of infants who heard coherent labels relative to infants who heard less coherent labels was because the coherent labels facilitated categorization and not because the less coherent labels inhibited categorization. These results are consistent with several studies that indicate that the representations infants form during statistical word segmentation have lexical status (Graf Estes et al., 2007; Hay, Pelucchi, Graf Estes, & Saffran, 2011; Saffran, 2001). Notably, the current results were found with 8-month-old infants, who are just beginning the task of mapping between words and their meanings. Despite their young age, the results indicate that infants are extracting and storing potentially meaningful units from speech.

Unlike previous studies using a similar paradigm, we did not find facilitation of category formation for completely novel linguistic items (e.g., Balaban & Waxman, 1997; Ferry et al., 2010; Fulkerson & Waxman, 2007). One possible explanation for this difference is that previous studies have typically included carrier phrases (e.g., "Look at the *toma*. Do you see the *toma*?") that may have made the task less challenging for infants. Carrier phrases sometimes facilitate infant performance in word learning (e.g., Fennell & Waxman, 2010). However, previous work has also shown that categorization facilitation occurs even when the carrier phrases are low-pass filtered and therefore unintelligible (Balaban & Waxman, 1997). Thus, the lack of carrier phrases in our study may not fully explain the difference between the current results and those of previous work.

A second methodological difference between this study and previous research lies in the duration of the familiarization phase. Previous studies using this general paradigm have varied in the length of the familiarization trials. In the current experiment, we used eight trials of 10 s each. However, most other studies used eight 20 s familiarization trials (e.g., Ferry et al., 2010; Fulkerson & Waxman, 2007; though see Balaban & Waxman, 1997 for a paradigm that included nine 10 s familiarization trials). In the present

study, our aim was to increase the difficulty of the task so that the dependent measure would have the sensitivity to detect a difference in performance between label conditions, given that previous studies have demonstrated successful categorization in the presence of novel labels. Thus, we used no carrier phrases and a brief presentation duration. These procedural differences were sufficient to prevent infants from categorizing in the presence of novel linguistic items (i.e., the unfamiliar labels in Experiment 2). Despite this increased difficulty in the current task, infants who heard statistically coherent labels were able to learn a novel category.

The difference in categorization performance as a function of label statistical structure indicates that the output of statistical learning shares properties of real words, consistent with theories that posit that statistical learning yields word-like representations (e.g., Saffran, 2001; Swingley, 2005). Note that this is not the only representational account that has been proposed. One alternative possibility is that rather than storing an integrated percept, infants might be storing chaining memories (e.g., when memorizing a sequence ABC, a chaining memory would consist of learning that A is linked to B and that B is linked to C; Endress & Mehler, 2009). Although the present findings do not rule this possibility out, both human data and computational modeling work indicate that learners are storing integrated chunks rather than chaining memories (Perruchet & Poulin-Charronnat, 2012). A second possibility is that infants might be storing bigram frequencies between syllables rather than an integrated unit. Again, while our current results do not rule this possibility out, it is unlikely for two reasons. First, prior explorations of statistical learning have demonstrated that learners detect non-adjacent relations, which cannot be accounted for via bigrams (e.g., Gómez, 2002; Newport & Aslin, 2004). Second, familiarity with sublexical bigrams (e.g., bisyllables contained with trisyllabic words) decreases as adult learners receive more exposure to artificial languages (e.g., Giroux & Rey, 2009); this is inconsistent with both bigram frequency and chaining memory accounts. Thus, our results are consistent with prior theoretical claims that statistical learning yields word-like representations, and further supports those claims by demonstrating that their word-likeness extends to facilitating categorization.

A further note relates to a potential distinction that could be made between a word-like unit and something that is merely familiar. It is not clear to us what it means to be truly word-like, above and beyond being a chunk that is highly familiar as a result of factors such as frequency and statistical coherence. By our view, this is essentially what it means to be a word form stored in an infant's proto-lexicon. We do not mean to imply that this is the end state of word learning, but rather it is an early point in the development of a lexicon (e.g., Mills, Plunkett, Prat, & Schafer, 2005). These stored items likely undergo further changes as a function of experience (e.g., Qiao & Forster, 2012). However, even early on, word forms that have been extracted from linguistic input are available to support further learning. For example, these word forms can be mapped to referents (e.g., Graf Estes et al., 2007), or used to support object categorization as in the present research.

We believe that the facilitation for categorization is in fact due to the fact that infants were more familiar with words than partwords (and unfamiliar labels) following statistical segmentation of the nonsense language. From our perspective, statistical segmentation of speech leads to the creation of a lexicon of stored word forms. This lexicon contains syllable clusters that are frequent and coherent; namely, clusters that co-occur reliably in the input (Giroux & Rey, 2009). These stored forms are available to be associated with referents. In the present study, words were more frequent and more statistically coherent than partwords (and unfamiliar labels, which did not occur at all). These factors together led infants to store words in memory, and treat them as familiar.

In contrast, the statistical characteristics of the partwords make it unlikely that infants formed or stored robust representations of them, even though they occurred repeatedly during segmentation. Specifically, the representation of each partword receives competition from both of the words from which it is formed. For example, in the phrase "statistical learning," parsing the word "statistical" makes parsing the word "stical" less likely. Data from behavioral experiments are consistent with this idea. As adult participants become more familiar with a whole word (like "statistical"), they become less familiar with the components of larger words (e.g., "stical") and partwords (e.g., "callearning") contained within the input (e.g., Giroux & Rey, 2009; Orbán et al., 2008). Therefore, even though infants were exposed to partwords in the input, their extraction of statistically coherent words may have inhibited the representation of the partwords to which they were also exposed. As such, the representations of the lexical items formed over the course of statistical learning should lead infants to treat words as more familiar than partwords. Because in the present study, words were more frequent than partwords, we cannot rule out the possibility that relative frequency contributed to the superior performance with words relative to partwords. However, prior research indicates that infants are learning about more than mere frequency in these statistical segmentation paradigms (e.g., Aslin, Saffran, & Newport, 1998; Thiessen & Saffran, 2003). Moreover, we believe that both of these statistical features of the environment—frequency and coherence—contribute to the formation of lexical representations.

Both frequency and statistical coherence are features that can help to transform a linguistic utterance into a good (i.e., familiar) candidate label (e.g., Colunga & Smith, 2002; Robinson & Sloutsky, 2007). From this perspective, the present results also have implications for theories of object categorization as well as theories of statistical learning. In particular, they highlight the importance of familiarity in transforming human utterances into labeling events. Research indicates that infants treat some symbols as better labels than others, and that the kinds of symbols they are willing to accept as labels narrows over the course of development (e.g., Ferry et al., 2013; Namy & Waxman, 1998; Woodward & Hoyne, 1999). Although infants are initially quite broad in what they will accept as plausible labels, this narrows considerably during the first years of life (Ferry et al., 2013; Woodward & Hoyne, 1999

Specifically, the kinds of symbols that older infants accept as potential labels tend to be familiar, and they have been systematically associated with categories in their environments (e.g., Colunga & Smith, 2002). Moreover, a body of research provides additional support for the importance of familiarity in the context of audiovisual associations (e.g., Robinson & Sloutsky, 2004). Auditory input can overshadow visual input and lead to decreased processing in the visual domain (e.g., Robinson & Sloutsky, 2004, 2007). However, when infants receive pre-familiarization with unfamiliar nonlinguistic sounds, the amount of auditory overshadowing is attenuated (Robinson & Sloutsky, 2007). These studies may shed light on the present results. Because partwords and unfamiliar labels were neither familiar to infants nor statistically coherent, they may have distracted infants from the visual stimuli in a way that coherent and familiar word items did not. This may have led to successful categorization for words, but not for the unfamiliar or partword labels. Thus, in agreement with previous findings, these results highlight the importance of both familiarity and statistical coherence in supporting infant category learning (e.g., Colunga & Smith, 2002; Robinson & Sloutsky, 2007).

These results should not be taken to mean that statistical learning is uniquely useful for creating proto-lexical representations. Acoustic cues to word segmentation (for example, the pauses around a monosyllabic utterance) likely allow infants to form the same kinds of representations, consistent with accounts suggesting that both acoustic and statistical segmentation share many processes in common (e.g., Thiessen et al., 2013). Rather, by providing evidence that the output of statistical learning is word-like, the present results support for the notion that statistical learning is a mechanism that underpins infants' ability to build a lexicon, and that statistical learning is related to real language functions and not simply an artifact of experimental laboratory paradigms. Because statistical learning allows infants to discover at least a few word forms, these word forms will be stored in memory during later opportunities to learn label-object or label-category mappings. The present findings are particularly remarkable because they show that young infants who have learned meanings for only a few native-language words (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999) are able to segment words via statistical learning and then use them to support further learning.

In summary, the present experiments showed that the representations formed during statistical learning influence object category formation early in infancy. These results dovetail with several studies that indicate that statistical learning produces word-like representations that can be used to support real linguistic tasks, such as label learning, in older infants and adults (Graf Estes et al., 2007; Mirman et al., 2008). Here, we found that young infants who are in the very early stages of word segmentation and word learning can use statistical learning to extract lexical units, then apply those units to support subsequent category learning. Infants can use statistical learning to solve the kinds of problems they face in their natural linguistic environments.

Acknowledgments

This research was supported by grants from the National Science Foundation to K.G.E. (BCS0847379) and to E.D.T. (BCS0642415). This material is based upon work supported by the National Science Foundation under Grant No. (0946825) awarded to L.C.E. We would like to thank Carolina Bastos, and the members of the Language Learning Lab at the University of California, Davis for their assistance with this research. We would also like to thank Marie Balaban, Alissa Ferry, Susan Hespos, and Sandra Waxman for sharing their stimuli, and David Rakison, Charles Kemp, Pierre Perruchet, and two anonymous reviewers for their comments on an earlier draft of this manuscript. We also thank the parents who generously contributed their time.

References

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

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology, 64*(1), 3–26.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253–3258.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*(2), B33–B44.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). *A new program for obtaining and organizing data in infant perception and cognition studies*, (Version 1.0). Austin: University of Texas.

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Erlbaum.

Colunga, E., & Smith, L. B. (2002). What makes a word. In *Proceedings of the annual conference of the cognitive science society* (Vol. 24, pp. 214–219).

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*(3), 351–367.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language and Hearing Research, 52*(2), 321.

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development, 81*(5), 1376–1383.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 1497–1510.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3-and 4-month-old infants: An advantage of words over tones. *Child Development, 81*(2), 472–479.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences, 110*(38), 15231–15235.

Fulkerson, A. L., & Haaf, R. A. (2003). The influence of labels, non-labeling sounds, and source of auditory input on 9- and 15-month-olds' object categorization. *Infancy, 4*, 349–369.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition, 150*, 218–228.

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science, 33*(2), 260–272.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436.
Graf Estes, K. (2012). Infants generalize representations of statistically segmented words. *Frontiers in Psychology, 3*.
Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words?: Statistical segmentation and word learning. *Psychological Science, 18*(3), 254–260.
Harris, Z. S. (1955). From phoneme to morpheme. *Language, 31*(2), 190–222.
Hay, J., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology, 63*, 93–106.
Hayes, J. R., & Clark, H. H. (1970). Experiments in the segmentation of an artificial speech analog. *Cognition and the Development of Language*, 221–234.
Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13*(2), 339–345.
Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3*(9), 323–328.
Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*(1), 1–23.
Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*(3–4), 159–207.
Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cognitive Development, 20*(1), 19–31.
Mirman, D., Magnuson, J. S., Graf Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*, 271–280.
Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development, 69*(2), 295–308.
Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127–162.
Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences, 105*(7), 2745–2750.
Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language, 66*(4), 807–818.
Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39*(2), 246–263.
Qiao, X., & Forster, K. I. (2012). Novel word lexicalization and the prime lexicality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75*(5), 1387–1401.
Robinson, C. W., & Sloutsky, V. M. (2007). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science, 10*, 734–740.
Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 906–914.
Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81*(2), 149–169.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52.
Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.
Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101–105.
Saffran, J. R., & Thiessen, E. D. (2007). Domain-general learning capacities. *Blackwell Handbook of Language Development*, 68–86.
Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*(1), 86–132.
Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience, 10*(1), 21.
Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*(1), 53–71.
Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*.
Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology, 39*(4), 706.
Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development, 3*(1), 73–100.
Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*(2), 172–175.
Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197–234.
Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development, 70*, 65–77.