COGNITIVE SCIENCE
A Multidisciplinary Journal

Tracking Multiple Statistics: Simultaneous Learning of Object Names and Categories in English and Mandarin Speakers

Chi-hsin Chen,[a] Lisa Gershkoff-Stowe,[b] Chih-Yi Wu,[c] Hintat Cheung,[d] Chen Yu[a]

[a]*Department of Psychological and Brain Sciences, Indiana University*
[b]*Department of Speech and Hearing Sciences, Indiana University*
[c]*Graduate Institute of Linguistics, National Taiwan University*
[d]*Department of Linguistics and Modern Language Studies, The Education University of Hong Kong*

Abstract

Two experiments were conducted to examine adult learners' ability to extract multiple statistics in simultaneously presented visual and auditory input. Experiment 1 used a cross-situational learning paradigm to test whether English speakers were able to use co-occurrences to learn word-to-object mappings and concurrently form object categories based on the commonalities across training stimuli. Experiment 2 replicated the first experiment and further examined whether speakers of Mandarin, a language in which final syllables of object names are more predictive of category membership than English, were able to learn words and form object categories when trained with the same type of structures. The results indicate that both groups of learners successfully extracted multiple levels of co-occurrence and used them to learn words and object categories simultaneously. However, marked individual differences in performance were also found, suggesting possible interference and competition in processing the two concurrent streams of regularities.

Keywords: Statistical learning; Word learning; Category learning; Simultaneous processing; Individual differences

1. Introduction

One of the more intractable puzzles in language development concerns how children learn the meaning of words. As Quine (1960) famously noted, the world presents a highly

Correspondence should be sent to Chi-hsin Chen, Department of Otolaryngology-Head and Neck Surgery, The Ohio State University, 915 Olentangy River Rd, Columbus, OH 43212. E-mail: chi-hsin.chen@osumc.edu

ambiguous learning context in which there are an infinite number of possible referents for any given word. For example, a child who encounters a tow truck for the first time must determine that the word *tow truck* refers to the object as a whole and not simply its color, its shape, the curious apparatus attached to its back, or to the salient looking construction worker standing beside it. How the child resolves the referential ambiguity problem has been the center of much debate. Researchers have variously argued that children solve Quine's problem by making use of conceptual biases (e.g., Markman, 1990; Markman & Wachtel, 1988), social cues (e.g., Baldwin, 1993; Tomasello & Akhtar, 1995), linguistic structures (e.g., Gleitman, 1990; Waxman & Booth, 2001), and statistical regularities (Smith & Yu, 2008; Yu & Smith, 2007).

To compound the problem, however, is the fact that there are multiple levels of ambiguity in natural word learning contexts. Children must learn not only the mapping between words and their meanings but also the relations between them. For example, the child who learns the word *dump truck* must also encode information about its relationship to the previously named *tow truck*, such as their similarities and differences (and possibly why these two objects have similar sounding names). Importantly, this ability is foundational to the establishment of categories of meaning. In this study, we investigate how statistical evidence over many word–referent pairs and across multiple learning trials can support the formation of object categories that may in turn be used to disambiguate the meaning of words.

There is considerable evidence to suggest that humans are equipped with powerful statistical learning capacities. Adults as well as infants and young children can detect and use regularities in their environment to solve a broad range of learning tasks. For example, infant learners can track patterns of distributional information to discover word boundaries (Saffran, Aslin, & Newport, 1996), phonetic categories (Maye, Werker, & Gerken, 2002), word-to-object mappings (Smith & Yu, 2008), and rudimentary syntax (Gómez & Gerken, 1999). In non-linguistic tasks, statistical learning has been demonstrated across a variety of situations, including the ability to make inferences (Xu & Denison, 2009), categorize objects (Plunkett, Hu, & Cohen, 2008), and process visual scenes (Fiser & Aslin, 2001). These and numerous other tasks suggest that human beings are remarkably skilled statistical learners.

The current literature in statistical learning has mainly focused on learning within a single task (e.g., speech segmentation *or* word learning). Although this approach has yielded considerable insight into the processes of learning and development, far less is known about statistical learning across simultaneously occurring tasks. Yet in everyday life, learners often encounter many different types of statistical regularities and perform multiple tasks at the same time. A few studies have shown that human learners are able to extract multiple levels of statistics embedded within a single set of stimuli. For example, Fiser and Aslin (2001) found that adults were able to learn multiple levels of regularities in complex visual scenes (see Fiser & Aslin, 2002, for an infant version). Romberg and Saffran (2013) found that adults were also able to concurrently track dependencies between adjacent and non-adjacent words in linguistic stimuli.

In addition to extracting a hierarchy of statistics within the same domain (e.g., visual *or* auditory), some recent studies have suggested that human learners were also capable of using statistical information to solve problems across domains. For example, in a study on simultaneous learning with stimuli from different domains, Yurovsky, Yu, and Smith (2012) showed that adults were able to use redundant information derived from a natural child-directed speech corpus to identify word boundaries in continuous speech and, at the same time, map newly segmented words to their correct referents (also see Cunillera, Laine, Càmara, & Rodríguez-Fornells, 2010; Shukla, White, & Aslin, 2011). Moreover, Thiessen (2010) observed that adult speech segmentation benefited when audio and visual information was presented simultaneously, in comparison to exposure to the audio stream alone. This result suggests that parallel processing of visual and audio information does not necessarily hinder learning in individual domains by increasing cognitive load. Instead, learning may be facilitated when multimodal information can be seamlessly integrated in real time.

Inspired by previous studies on concurrent statistical learning, the overall aim of this study is to examine how human learners extract multiple levels of statistical regularities from visual and speech stimuli. With this general goal in mind, this study focuses on a specific learning task—using statistical information to learn both object names and object categories. Numerous studies have shown a close link between word learning and object categorization. Knowing which objects are members of the same category helps learners better understand the meaning of category labels (e.g., Xu & Tenenbaum, 2007). In turn, knowing that some objects share the same labels facilitates comparisons of the objects and guides learners' attention to the similarities they share (e.g., Colunga & Smith, 2005; Sloutsky, Lo, & Fisher, 2001). However, it remains unclear whether such links between word learning and object categorization can be established *de novo* through a single training session in which learners need to learn the mappings between words and referents and simultaneously form object categories.

This study investigates the topic by using a cross-situational learning paradigm in which participants are exposed to multiple learning trials, each consisting of multiple words and multiple objects. Within an individual trial, information is ambiguous about which words map onto which objects. Thus, participants are required to keep track of co-occurrences between the two. Numerous studies have demonstrated that human learners are capable of tracking co-occurring information across multiple learning trials to disambiguate between consistent and inconsistent co-occurrences and ultimately build correct word–referent mappings (e.g., Kachergis, Yu, & Shiffrin, 2013; Smith, Smith, & Blythe, 2011; Suanda & Namy, 2012; Yu & Smith, 2007; Yurovsky, Yu, & Smith, 2013). This study expands upon previous research by testing whether adult learners can extract multiple levels of statistical regularities to learn object names and categories concurrently.

The experiments reported below use a modified cross-situational word learning design that includes two levels of regularities: word-to-object mappings and syllable-to-category associations. In each training trial, participants see four objects and hear four words presented in random order. However, they are not provided with information about which

object is matched to which word. To find the correct mappings, participants have to keep track of the co-occurrence regularities between words and objects across different trials. In addition to word–object co-occurrences, another level of regularity is embedded in the training stimuli. Objects belonging to the same category have a similar looking object part and are mapped to words either starting with the same syllable or ending with the same syllable.

One hypothesis is that whether learners form syllable-to-category associations partially depends on how well they learn individual word-to-object mappings. If this is the case, then learners need at least a few instances to extract commonalities among category members. Given the simultaneous presentation of two streams of statistical regularities in this study (i.e., individual object level vs. category level), having high sensitivity to category-relevant features may interfere with word learning because of competition for attention and processing. Learners who fail to overcome this interference may have low word learning performance, which can then hinder the formation of correct syllable-to-category associations. On the other hand, with sufficient learned word–object pairs available, having high sensitivity to category-relevant features may allow learners to easily detect the syllable-to-category associations. In the latter case, we might see not only successful learning of word-to-object mappings and syllable-to-category associations, but also bootstrapping between these two processes.

Following the general ideas described above, we designed and implemented two experiments to examine concurrent statistical learning of word-to-object mappings and object categories in two different contexts and with speakers of two different languages. In Experiment 1, we encoded category-indicating information in either the initial syllables (Same-Initial-Syllable structure, e.g., *joti* and *josen*) or the final syllables (Same-Final-Syllable structure, e.g., *joti* and *feti*) of English-like pseudo-words. The Same-Initial-Syllable structure is analogous to the real-world adjectives in adjective–noun phrases (e.g., the word *red* in *red car* and *red book*) or the modifiers in compound nouns (e.g., the word *cheese* in *cheesecake* and *cheese stick*), whereas the Same-Final-Syllable structure simulates the head nouns in adjective–noun phrases (e.g., the word *car* in *red car* and *blue car*) or compound nouns (e.g., the word *cake* in *cheesecake* and *cupcake*). The goal of this first experiment was to test whether native English speakers were able to learn both word-to-object mappings and syllable-to-category associations through a single training session, and whether the learning outcome was robust across the two different linguistic structures. In Experiment 2, a group of native Mandarin speakers was exposed to Mandarin-like pseudo-words with the same general training structures used in Experiment 1. In modern Mandarin, object names are typically disyllabic compounds. Moreover, objects belonging to a category often share a final syllable in their labels. Therefore, testing Mandarin speakers allowed us to answer two questions: (a) whether Mandarin speakers are also able to simultaneously learn two levels of regularities in the stimuli, and (b) if so, how prior experiences (e.g., differences in the prevalence of compound nouns in English and Mandarin differentially) affect learners' sensitivity to the properties in the training stimuli.

2. Experiment 1

Experiment 1 investigated adult English speakers' ability to extract multiple levels of co-occurrence statistics not only to learn word-to-object mappings but also to form object categories.

2.1. Participants

Participants were 72 undergraduate native speakers of English (39 women, M_{age}: 19.4) at Indiana University who received course credit for volunteering.

2.2. Design and stimuli

There was a total number of 18 target objects in the training phase divided into three categories, with six items in each category. Members within a category had an attached part that looked similar. In addition, objects belonging to the same category were mapped to disyllabic pseudo-words that shared one syllable with each other. Over the training procedure, each word–object pair occurred 12 times, yielding a total of 54 trials (3 categories × 6 pairs × 12 repetitions/4 pairs per trial). Each trial took 12 s. The entire training session lasted 10.8 min.

As shown in Fig. 1A, each trial contains four novel objects and four pseudo-words. Within a single trial, learners do not have the information to determine which word goes with which object. For example, there is no information to identify which object is mapped to the word *josen* in either Trial 1 or Trial 2 alone. Nonetheless, if participants remember hearing the word *josen* and seeing the triangular object, then they should be able to infer that *josen* is mapped to this object in Trial 2. That is, across learning trials, each pseudo-word co-occurs consistently with only one object. Also shown in Fig. 1B, there are three to-be-learned syllable-to-category mappings. For example, the words starting with the syllable *jo-* are mapped to objects with a similar looking spiral tail, whereas the words starting with the syllable *che-* are mapped to objects with a different part. Critically, participants were asked to learn only word-to-object mappings, but not the higher-order regularities between the linguistic features of labels and the visual features shared by members of the same object category.

Half of the participants were trained with the Same-Initial-Syllable structure in which words in a category began with the same syllable (e.g., *joti* and *josen*), whereas the other half were trained with the Same-Final-Syllable structure in which words in a category ended with the same syllable (e.g., *joti* and *feti*). To control for item effects, we used two sets of stimuli for each structure. The stimuli in Sets Eng-A and Eng-B had words in a category starting with the same initial syllable, whereas Sets Eng-C and Eng-D had words in a category ending with the same final syllable. The auditory input was based on natural human voices generated by AT&T Natural Voices. Each syllable was generated separately and modified to have comparable loudness and matching lengths (350 ms). The

(A)

Trial 1   Trial 2

josen ... chedi ... mooli ... chetow   chepa ... josen ... mooray ... joti

(B)

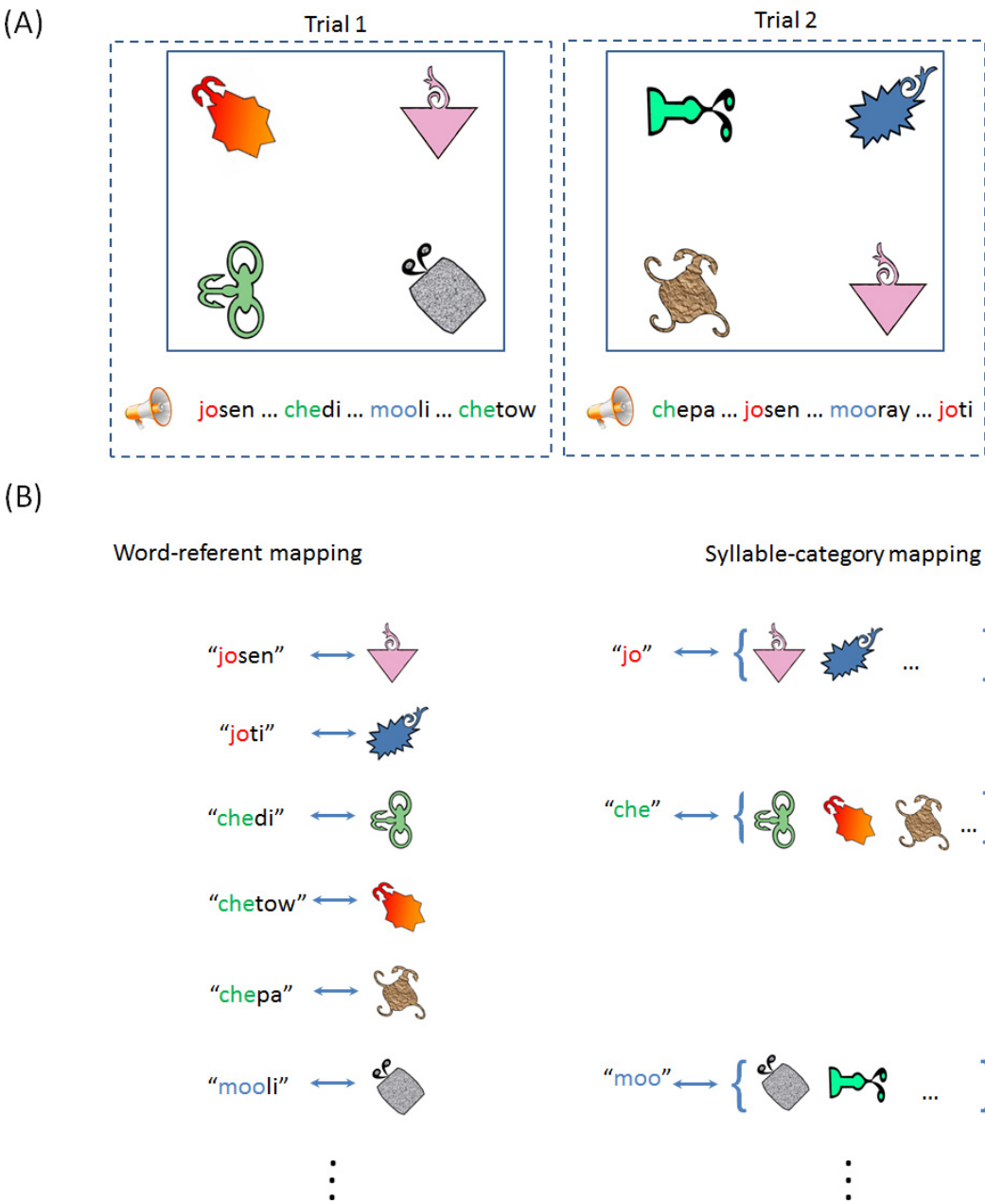Word-referent mapping   Syllable-category mapping

Fig. 1. (A) Examples of training trials. Learners had to use the co-occurrence information across trials to find the correct word-to-object mappings and syllable-to-category associations. (B) Correct word–object mappings and syllable-category mappings embedded in Trials 1 and 2.

pseudo-words were then created by connecting discrete syllables together. Because of matched loudness and lengths, each syllable in the concatenated pseudo-words was not specifically stressed.

2.3. Procedure

Before the training session began, participants were informed that they would go through a series of learning trials where they saw four objects and heard four words within each trial. They were explicitly told that the order of the words presented in each trial was not related to the locations of the objects on the screen and that their goal was to learn the correct word-to-object mappings by the end of the training session.

During training, participants were presented with 54 slides, each containing four objects and four auditory labels. The labels in each trial were presented in a random order. Following training, participants were tested with two types of tests:

1 Mapping: The 18 mapping trials assessed how well participants learned the names of the training objects. In that task, participants heard one trained word at a time and had to select its referent from four trained objects.
2 Generalization: The nine generalization trials assessed how well participants learned the syllable-to-category associations. Participants heard one novel word on each trial and were asked to select its referent from three novel objects, each containing the object part that corresponded to the particular feature of one category. For example, participants might hear a novel label *jorow*, which had the same initial syllable as the words matched to the training objects with a spiral part seen in Fig. 1B. Participants had to select its referent from three novel objects, one of which had the spiral part. If participants chose that object, this indicated that they had acquired the relationships between the initial syllables of the labels and the corresponding perceptual features of the categories. Each category was tested three times.

2.4. Results

We first report preliminary analyses that check for a stimulus set effect within the Same-Initial-Syllable and Same-Final-Syllable structures. The following two sets of analyses examine whether adult English speakers were able to learn word-to-object mappings and syllable-to-category associations. Thereafter, we take a closer look at the relationship between learners' word and category learning.

2.4.1. Preliminary analyses

Preliminary analyses were conducted to determine whether different sets of stimuli within each language affected participants' learning of the word-to-object pairings. There was no accuracy difference between participants trained with Sets Eng-A and Eng-B for the Same-Initial-Syllable structure ($t(34) = 0.660$, *n.s.*), nor was there a difference between participants trained with Sets Eng-C and Eng-D for the Same-Final-Syllable

structure ($t(34) = -0.432$, *n.s.*). Because different training sets did not affect learning, all subsequent analyses were collapsed across training sets for each structure.

2.4.2. Learning word-to-object mappings

The first key question was whether participants learned individual word-to-object mappings. As illustrated on the left-hand side of Fig. 2, learners in both Same-Initial-Syllable and Same-Final-Syllable conditions, on average, learned at least half of the word–object pairs. Chance level analyses confirmed that the two groups both performed significantly above chance (i.e., 25%, as learners had to pick the target from four alternatives) in the Mapping task (Same-Initial: $M = .6049$, $t(35) = 7.655$; Same-Final: $M = .5401$, $t(35) = 6.846$, *ps* < .001). In addition, both groups had comparable word learning accuracies ($t(70) = 1.032$, *n.s.*). These results indicate that, regardless of the training structures, English speakers were able to use co-occurrences to form word-to-object mappings after 10.8 min of training.

Because the stimuli contained two levels of regularities, word-to-object mappings and syllable-to-part associations, one could argue that another way to succeed in the Mapping task is to use the initial or final syllable of a test word as a cue and pick one object containing a specific part as the target referent without necessarily knowing individual word-to-object mappings. To test this possibility, we examined participants' performance on Mapping trials containing distractors from the same category as the target referents (subsequently termed Same-Category distractors). Of the 18 Mapping trials, seven contained one Same-Category distractor. If participants relied solely on the syllable-to-part associations without knowing individual word-to-object mappings, the presence of Same-Category distractors should affect their accuracy. We first tested whether the mean accuracies
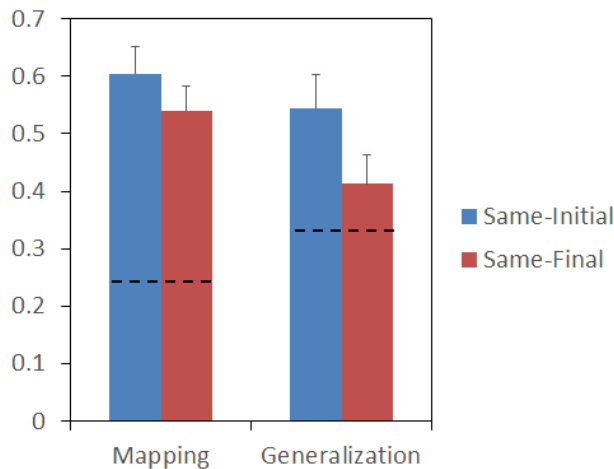

Fig. 2. Mean proportion of accurate responses (and standard errors of the means) in the Mapping and Generalization tasks in Experiment 1. Learners had to pick one object from four alternatives in the Mapping task. Therefore, the chance level was one fourth. Learners heard one novel word and had to choose its referent from three novel objects in the Generalization task. Therefore, the chance level was one third.

of trials containing a Same-Category distractor were above chance and then compared the mean accuracies of trials with and without a Same-Category distractor. Learners in the Same-Initial-Syllable condition had above-chance Mapping performance when tested with both trials with and trials without a Same-Category distractor (With: $M = .5754$, $t(35) = 6.810$; Without: $M = .6237$, $t(35) = 7.406$, ps < .001). Accuracies for these two types of trials were not significantly different from each other ($t(35) = -1.373$, *n.s.*). Similarly, participants in the Same-Final-Syllable condition also performed above chance for both types of test trials (With: $M = .5000$, $t(35) = 4.958$; Without: $M = .5808$, $t(35) = 7.126$, ps < .001). Their performance difference for these two types of trials was also not significant ($t(35) = -1.532$, *n.s.*). Together, these results confirm that participants were able to accurately pick the target object when hearing a word, even when the test trials contained a distractor that shared perceptual similarity (i.e., part) with the target.

2.4.3. Learning syllable-to-category associations

The second key question was whether learners could successfully extract the syllable-to-category associations from the training stimuli. As a group, learners trained with the Same-Initial-Syllable structure performed significantly above chance (i.e., 33.3%, as learners had to pick the target from three alternatives), suggesting that they reliably used the initial syllables of labels as a cue in the Generalization task (Same-Initial: $M = .5432$, $t(35) = 3.576$, $p = .001$). There was also a marginally significant trend suggesting that learners trained with the Same-Final-Syllable structure also used final syllables of labels in the Generalization task (Same-Final: $M = .4136$, $t(35) = 1.959$, $p = .058$). Moreover, although mean accuracy was numerically greater for the Same-Initial-Syllable group, the performance difference of these two groups was not statistically significant ($t(70) = 1.811$, $p = .074$).

2.4.4. Concurrent word and category learning

The third set of analyses examined whether learners' overall performance in the Generalization task was associated with their Mapping performance (Fig. 3). Significant positive correlations between learners' overall Mapping and Generalization scores were found for both groups (Same-Initial: Pearson's $r = .716$, $p < .001$; Same-Final: Pearson's $r = .503$, $p < .01$), suggesting that the more word–object pairs participants learned, the better their Generalization performance.

These analyses were based on the total number of correct answers in the Mapping and Generalization tasks, aggregated across all three word–object categories. However, learners who had the same total number of correct answers in the Mapping and the Generalization tasks might reveal very different learning profiles. For example, two participants, A and B, who both have six correct answers in the Mapping task and three correct answers in the Generalization task, may nonetheless show very different patterns of learning. At one extreme, Participant A might learn all six word–object pairs from a single category and correctly pick all three novel objects from that specific category in the Generalization task but learn nothing about the other two categories. At the other extreme, Participant B might learn all six word–object pairs from one category without noticing
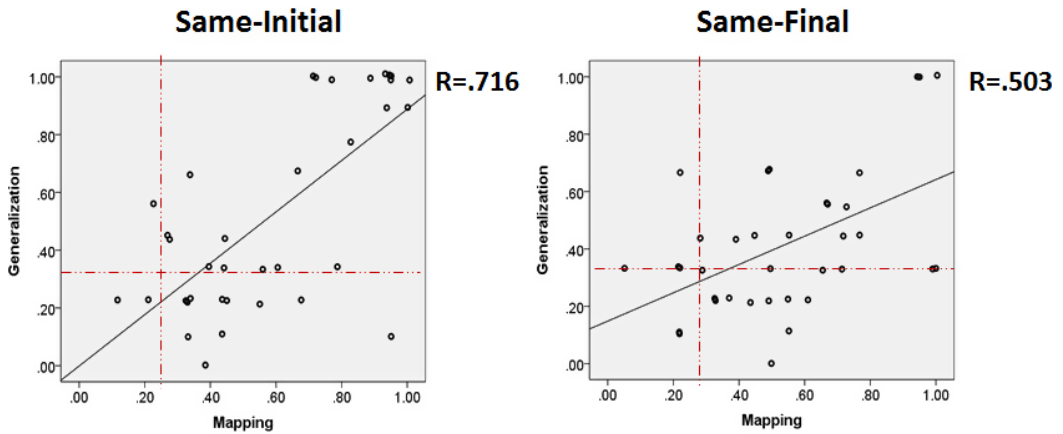
Fig. 3. Correlations between Mapping and Generalization performance in Experiment 1. The solid line in each graph represents the regression line. The vertical dash-dot lines indicate the chance level in the Mapping task (i.e., 25%), whereas the horizontal dash-dot lines indicate chance in the Generalization task (i.e., 33.3%).

the syllable-to-category associations. That participant might then randomly pick objects in the Generalization trials and, accidentally, have three successful shots, one for each category. If we look only at Participants A and B's overall Mapping and Generalization performance, the data would reveal that one third of the trials were correct in each task. However, Participant A's learning profile indicates a tight relation between Mapping and Generalization, whereas Participant B's performance suggests two separate learning processes. To address this issue, we next analyzed the data at the category level to examine whether the more word–object pairs participants learned for a category, the more likely they were to detect the syllable-to-category association for that specific category.

We used a generalized estimating equations (GEE) method to examine whether learners' Mapping scores for a category was a good predictor of their Generalization score for that specific category and, vice versa, whether their Generalization performance predicted their Mapping. The GEE method was selected because traditional linear models assume independence among data points. However, in our design, each participant contributed three data points, one for each word–object category. Therefore, we used the GEE method to account for correlations among observations from the same participant (Liang & Zeger, 1986). We first used the number of correct Mapping trials for each category as the predictor and the number of correct Generalization answers for the same category as the dependent variable. In both Same-Initial-Syllable and Same-Final-Syllable conditions, learners' Mapping scores for a category was a significant predictor for their Generalization scores for the same category (Same-Initial: Wald χ^2 = 18.348, p < .001; Same-Final: Wald χ^2 = 9.142, p < .01). In addition, learners' Generalization scores for a category predicted their Mapping scores as well (Same-Initial: Wald χ^2 = 21.119, p < .001; Same-Final: Wald χ^2 = 7.393, p < .01). These results suggest that the correlations between learners' aggregated Mapping and Generalization scores are driven by the links between word and category learning for individual categories. The more word–object pairs

participants learned for a category, the more likely they used the initial or final syllable of a label as a cue in the Generalization task.

2.5. Discussion

Consistent with previous studies on cross-situational learning, participants in this study were able to rely on the co-occurrence information between labels and objects to build individual word–referent mappings. This was true regardless of the word-to-referent training structure. Even though the labels of objects within each category were fairly similar to each other (i.e., words in a category began or ended with the same syllable), participants were able to distinguish between the words and perform above chance in the Mapping task after only 10.8 min of training.

Importantly, despite the fact that learners were only instructed to find the mappings between words and objects and not informed of the category structures, those trained in the Same-Initial-Syllable condition, as a group, performed significantly above chance in the Generalization task. This result indicates that adult learners were able to extract the syllable-to-category associations without any instruction to pay attention to the higher-level regularities in the stimuli. Moreover, correlational analyses showed that the more word–object pairs participants learned, the more likely they were to notice the syllable-to-category associations and use the shared syllable as a cue in the Generalization task.

There was a trend suggesting that participants trained with the Same-Initial-Syllable structure had higher Generalization scores than their counterparts trained with the Same-Final-Syllable structure. Furthermore, the mean Generalization performance of learners in the Same-Final-Syllable condition did not differ from chance, indicating that Generalization was not as robust in this group. One possible reason for this result is related to the informativeness or psychological saliency of initial and final syllables in English words. Kessler and Treiman (1997) suggested that because of the relative unpredictable onset-vowel, but predictable vowel-coda associations in English syllable structure, the beginnings of words are generally more informative in both word recognition and production. In addition, due to the fact that the majority of English words begin with a stressed syllable, it has been argued that the beginning portion of a word is more psychologically salient than other parts and thus may draw more attention than the final portion of the word (e.g., Cutler & Carter, 1987). Even though our stimuli were created by concatenating independently generated syllables together and each syllable was not specifically stressed, it is possible that the psychological saliency or informativeness of the initial syllables in real English words led to learners' heightened attention toward initial syllables and facilitated category detection in the Same-Initial-Syllable condition. Hupp, Sloutsky, and Culicover (2009) found that when adult speakers of English were presented with a two-syllable nonsense target word (e.g., ta-te) and asked to judge whether a "pre-changed" item ("*be*-ta-te") or a "post-changed" item ("ta-te-*be*") was more similar to the target, participants showed a preference for the post-changed item. Although not addressed directly in that study, this result implies that English speakers might take words with the same beginnings as more similar than words with the same endings. If this interpretation

is true, the heightened (psychological) similarity of words sharing the initial syllables in the Same-Initial-Syllable condition may have facilitated the discovery of syllable-to-category associations and led to better Generalization performance.

The above explanation implies that, due to the structure in the ambient language environment, English speakers may not have paid sufficient attention to the final syllables of the novel pseudo-words and thus failed to perform above-chance Generalization in the Same-Final-Syllable condition. Following this logic, one would predict that speakers of a language that potentially tunes their attention more to the final syllables of words than English does should show robust learning of the Same-Final-Syllable structure in the current design. One such language is Mandarin. Experiment 2 was designed to replicate the current findings that adult learners are able to use co-occurrence information to learn word-to-object mappings and syllable-to-category associations simultaneously. Moreover, we were interested in testing whether speakers of Mandarin, a language in which final syllables of object names are often associated with object category membership, are able to perform above chance when trained with the Same-Final-Syllable structure.

3. Experiment 2

In the Mandarin language, compounding plays a major role in the formation of words (Li & Thompson, 1981). By some estimates, over 70% of modern Mandarin words are compound words (Tsai, Lee, Lin, Tzeng, & Hung, 2006; Zhou, Marslen-Wilson, Taft, & Shu, 1999); of those, approximately 29% link two nouns together to form nominal compounds (Huang, 1998). In contrast, far fewer compounds are found in English. Data extracted from the CELEX database suggest that less than 2.8% of all English word types are noun–noun compounds (Janssen, Bi, & Caramazza, 2008; Plag, Kunter, & Lappe, 2007). In modern Mandarin, object names are typically disyllabic compounds. Items belonging to the same category tend to share a final syllable signifying the name of the category. For example, in English, the words "pork," "beef," and "chicken" do not share any constituent in their forms. In Mandarin, however, they share a final syllable, *rou4*[1] (meat). The second syllable of the words *zhu1-rou4* (pork), *niu2-rou4* (beef), and *ji1-rou4* (chicken) indicates membership in the category of "meat." Given that words ending with the same syllable are likely to belong to the same category, one learning strategy for Mandarin speakers is to form categories based on the final syllables of the labels.[2] Therefore, the first goal of Experiment 2 was to test whether Mandarin speakers were able simultaneously to learn word-to-object mappings and to extract the syllable-to-category structure in the Same-Final-Syllable condition.

Like English, adjectives in a noun phrase in Mandarin come before the head noun, as in *red car* (Li & Thompson, 1981).[3] For example, the initial syllable in the phrases, *hong2-che1* (red car), *hong2-mao4* (red hat), and *hong2-shu1* (red book), indicates that each object shares the same color red. With this common noun phrase structure in the language, we expect that Mandarin speakers should be able to learn the syllable-to-category associations in the Same-Initial-Syllable condition as well. The second goal of this

experiment was to replicate the findings in Experiment 1 and test whether Mandarin speakers were able to simultaneously learn word-to-object mappings and syllable-to-category associations in the Same-Initial-Syllable condition.

3.1. Participants

Participants were 48 native speakers of Mandarin (23 women, M_{age}: 20.5) recruited from National Taiwan University, Taiwan. Participants were enrolled in a course designed for students who did not pass the first stage of the High Intermediate level test of the General English Proficiency Test (GEPT),[4] a standardized English proficiency test used in Taiwan, and who scored low on self-reports. Thus, although participants in Experiment 2 had some exposure to English, none was a proficient English–Mandarin bilingual.

3.2. Design and stimuli

The visual stimuli in Experiment 2 were identical to those in Experiment 1. Novel pseudo-words that followed the phonological rules of Mandarin were created. These Mandarin-like pseudo-words were closely matched to the stimuli used in Experiment 1 in their phonetic features. For example, the English-like pseudo-word *feti* has its counterpart *fe3-ti2* in the Mandarin pseudo-word set. In other words, the Mandarin-like pseudo-words were considered the Mandarin version of the English stimuli, with additional tonal information. It is important to note that the tonal information did not serve any contrastive function in this study. From an information processing perspective, the tonal information can be viewed as an additional but non-diagnostic dimension in this task.

As in Experiment 1, half of the Mandarin speakers were trained with the Same-Initial-Syllable structure and half trained with the Same-Final-Syllable structure. Two sets of words were used for each structure to control for item effects, Sets Mand-A and Mand-B had words in a category starting with the Same-Initial-Syllable while Sets Mand-C and Mand-D had words ending with Same-Final-Syllable.

3.3. Procedure

The procedure was identical to that of Experiment 1, except that the instruction and auditory stimuli were presented in Mandarin.

3.4. Results

We first report preliminary analyses that check for the presence of stimuli set effect within each structure. Following that, we examine whether Mandarin speakers were able to learn word-to-object mappings and syllable-to-category associations. We then analyze the relationship between their Mapping and Generalization performance. In the final set of analyses, we test whether English and Mandarin speakers performed differently from each other.

3.4.1. Preliminary analyses

Preliminary analyses revealed no difference in accuracy between participants trained with Sets Mand-A and Mand-B ($t(22) = 1.067$, *n.s*), nor was there a difference in accuracy between participants trained with Sets Mand-C and Mand-D ($t(22) = 1.329$, *n.s.*). All subsequent analyses were thus collapsed across training sets for each structure.

3.4.2. Learning word-to-object mappings

Our first set of analyses tested whether Mandarin speakers, like their English-speaking counterparts, successfully learned word-to-object mappings. Chance analyses revealed that Mandarin speakers trained with both Same-Initial-Syllable and Same-Final-Syllable structures succeeded in learning individual word-to-object mappings (left-hand side of Fig. 4, Same-Initial: $M = .5833$, $t(23) = 5.894$; Same-Final: $M = .500$, $t(23) = 4.153$, $ps < .001$). These two groups did not perform differently from each other in the Mapping task ($t(46) = 1.009$, *n.s.*). The results replicate the findings of Experiment 1 and further support the idea that adult learners are skilled at using co-occurrences to learn word-to-object mappings.

The second set of analyses examined whether Mandarin speakers reliably selected the target referents in trials with and without a Same-Category distractor. Participants in the Same-Initial-Syllable condition had above-chance accuracies for both types of trials (With: $M = .5298$, $t(23) = 4.219$; Without: $M = .6174$, $t(23) = 6.462$, $ps < .001$). Although they had numerically higher performance for trials without a Same-Category distractor, the difference did not reach statistical significance ($t(23) = 1.935$, $p = .065$). Participants trained in the Same-Final-Syllable condition also were able to reliably select the target referents in both types of test trials (With: $M = .4048$, $t(23) = 2.311$, $p < .05$; Without: $M = .5606$, $t(23) = 5.004$, $p < .001$). However, they had significantly better performance for trials without a Same-Category distractor ($t(23) = 3.483$, $p < .01$).
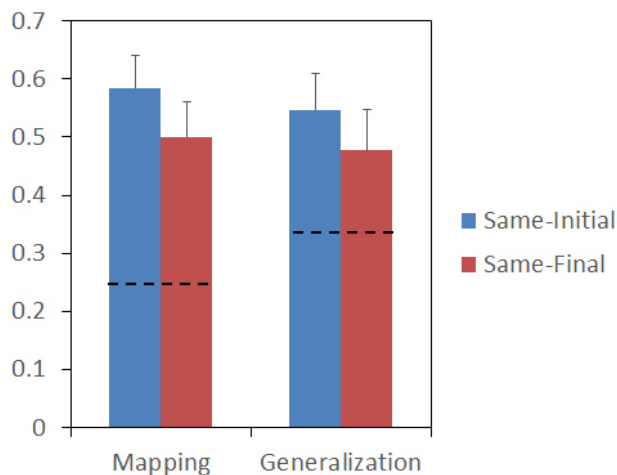

Fig. 4. Mean proportion of accurate responses (and standard errors of the means) in the Mapping and Generalization tasks in Experiment 2.

We next analyzed the error patterns for trials containing a Same-Category distractor. Three participants were excluded from this analysis because of perfect performance on those trials. Those who erred on the Same-Category distractor trials were significantly more likely to select the Same-Category distractor than expected by chance (chance = 1 of 3 distractors; $M = .4837$, $t(20) = 2.102$, $p < .05$). Taken together, these results indicate that the Mandarin speakers, overall, were able to successfully pick the target referent of a word in both trials with and without a Same-Category distractor. However, Mandarin speakers in the Same-Final-Syllable condition were significantly affected by Same-Category distractors. Critically, when they made errors on those trials, they tended to mistakenly select the distractor that had a similar object part as the target rather than objects from other categories. This result suggests that (partial) knowledge of the syllable-to-object part associations interfered with participants' performance in the Mapping task.

3.4.3. Learning syllable-to-category associations

We next asked whether Mandarin speakers successfully formed syllable-to-category associations (right-hand side of Fig. 4). Learners in both Same-Initial-Syllable and Same-Final-Syllable conditions performed significantly above chance (Same-Initial: $M = .546$, $t(23) = 3.377$, $p < .01$; Same-Final: $M = .4861$, $t(23) = 2.132$, $p < .05$); they also did not perform differently from each other in the Generalization task ($t(46) = .728$, *n.s.*). These results suggest that Mandarin speakers reliably used the initial or final syllable of a novel label as a cue in the Generalization task.

3.4.4. Concurrent word and category learning

As in Experiment 1, we also found positive correlations between Mandarin speakers' overall Mapping and Generalization scores (Fig. 5: Same-Initial: Pearson's $r = .439$, $p < .05$; Same-Final: Pearson's $r = .888$, $ps < .001$). The more word–object pairs participants learned, the more likely they were to notice the syllable-to-category associations. As before, GEE methods were used to examine whether correlations between aggregated Mapping and Generalization scores resulted from correlations within individual categories. Consistent with previous findings, the results showed that learners' Mapping performance for a category was a significant predictor of their Generalization score for the same category (Same-Initial: Wald $\chi^2 = 7.904$, $p < .01$; Same-Final: Wald $\chi^2 = 78.804$, $p < .001$). In addition, their Generalization scores for a category predicted their Mapping performance as well (Same-Initial: Wald $\chi^2 = 8.778$, $p < .01$; Same-Final: Wald $\chi^2 = 54.370$, $p < .001$).

3.4.5. Cross-experiment comparisons

The last set of analyses tested whether English and Mandarin speakers performed differently from each other. We conducted two separate 2 (Speaker: English vs. Mandarin) × 2 (Structure: Same-Initial vs. Same-Final) analyses of variance, one with participants' Mapping scores as the dependent measure and the other with their Generalization scores. There were no significant main effects or interactions in any of the comparisons, $ps > .05$. The results suggest that even though nuanced manipulations (e.g.,
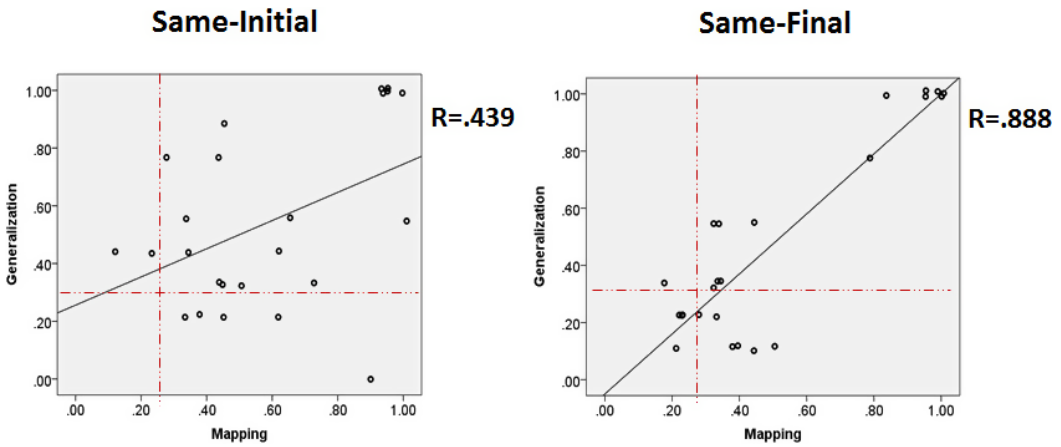
Fig. 5. Correlations between Mapping and Generalization performance in Experiment 2.

locations of shared syllables) differentially affected speakers of different languages, our general findings about adults' Mapping and Generalization were reliable across speaker groups and across conditions.

3.5. Discussion

As in Experiment 1, Mandarin speakers were able to use co-occurrence information to learn word-to-object mappings. Even though they had above-chance performance for both Mapping trials with and without a Same-Category distractor, Mandarin speakers, particularly those in the Same-Final-Syllable condition, tended to be affected by the presence of Same-Category distractors. This provides indirect evidence of their (partial) category knowledge. More direct evidence of their category learning comes from their Generalization performance. Consistent with our prediction, Mandarin speakers noticed the syllable-to-category associations in both Same-Initial-Syllable and Same-Final-Syllable structures and performed above chance in the Generalization task. These results together suggest that adult Mandarin speakers are able to use co-occurrences to simultaneously learn word-to-object mappings and syllable-to-category associations.

Similar to the patterns seen in Experiment 1, there were positive correlations between participants' Mapping and Generalization performance. The more word–object pairs they learned, the more likely they were able to use the phonological features of labels in categorizing novel objects. This makes sense given that learners need at least a few correct word–object pairs to extract commonalities across objects and across words before they are able to infer the relationship between the phonological features of labels and the visual features of objects. In general, learners who succeeded in learning word-to-object mappings excelled in category learning, whereas those having low Mapping scores tended to have low Generalization scores. Yet the scatterplots in Fig. 5 (as well as those in Fig. 3) show that some participants had fairly high Mapping scores but low

Generalization scores. These learners may have focused on the word learning task they were instructed to do and either deliberately ignored the higher-order syllable-to-category associations or simply did not have the processing resources to successfully extract this stream of regularity. On the other hand, some learners had low Mapping but high Generalization scores. These learners may have noticed the syllable-to-category associations but did not remember which word was mapped to which specific object during the testing phase. This result seems to contradict our assumption that learners need a few word–object pairs before forming syllable-to-category associations and suggests that some learners may have been able to bypass the word-to-object mapping step. Theoretically, tracking the syllable-to-object-part co-occurrences *alone* does allow learners to form syllable-to-category associations. This can be accomplished by tracking the correspondences between the number of words containing a certain syllable and the number of objects containing a specific part across trials. For example, if learners notice that trials containing two *jo*-words always contain two objects with a specific tail while trials containing one *jo*- word only have one object with that specific part, then they can potentially form syllable-to-part associations without learning individual word-to-object mappings. However, tracking the number correspondences between syllables and object parts is less straightforward than tracking word-to-object mappings. To assess the possibility that learners formed syllable-to-category associations *solely* by tracking the number correspondences between syllables and object parts, we took a closer look at the data of participants having a low to fair Mapping score (i.e., below 45% correct) but a high Generalization score (i.e., above 70% correct) and examined their performance across different categories. The results showed that the participants who produced this low–high pattern all had one category with a perfect or near-perfect Mapping score (at least 5 of 6 trials correct) and a perfect Generalization score (3 of 3 correct) for that category. It is likely that these learners may have noticed the syllable-to-category association for the well-learned category and then applied this knowledge to categories with low Mapping scores.

These different learning patterns suggest that when presented with two concurrent streams of statistical regularities, individual learners approach them differently. Some learners not only succeeded in the instructed word-to-object Mapping task but also extracted the syllable-to-category associations, even without any instruction to do so. Many learners focused on one task, be it the instructed word learning task or the uninstructed category learning task, but failed the other task. And still others failed to extract either stream of regularity.

4. General discussion

Two experiments were conducted to examine adult learners' ability to extract multiple statistics in simultaneously presented visual and auditory input. Experiment 1 tested whether English speakers were able to use co-occurrences to learn word-to-object mappings and concurrently form object categories based on the commonalities across training stimuli. Experiment 2 replicated the first experiment and further examined whether

speakers of Mandarin, a language in which final syllables are more predictive of category membership than English, were able to learn words and form object categories when trained with the same type of structures.

4.1. Multiple levels of statistical regularities

Consistent with the findings from previous cross-situational word learning studies (e.g., Yu & Smith, 2007), both English and Mandarin speakers in this study were able to use co-occurrence regularities to learn word-to-object mappings. To our knowledge, this study is the first to test speakers from two typologically different languages on their word learning performance in the same cross-situational learning paradigm. Overall, comparable Mapping performance from these two groups of participants suggests that the findings are reliable across languages.

In addition, many of the participants were able to form syllable-to-category associations and consistently used this cue in Generalization. This indicates that adult learners are able to extract multiple levels of regularities cross-modally and use them to simultaneously perform word and category learning tasks. Learners accomplish these tasks by matching co-occurring auditory stimuli to visual objects, extracting the phonological patterns across multiple words and the visual patterns across multiple objects, and associating these phonological patterns with the visual patterns. This study extends previous findings (e.g., Fiser & Aslin, 2001; Romberg & Saffran, 2013) by showing that adult learners are able to extract regularities in both visual and auditory modes simultaneously and form cross-modal associations at both word-to-object and syllable-to-category levels.

One question raised by these findings is whether word and category learning processes compete with each other. Compared to previous cross-situational word learning studies (e.g., Suanda & Namy, 2012; Yu & Smith, 2007), the current findings indicate that adding the category learning component did, in fact, interfere with successful word learning. In this study, each word–object pair co-occurred 12 times across training, resulting in twice as many training trials as in several previous cross-situational word learning studies (e.g., Suanda & Namy, 2012; Yu & Smith, 2007). However, learners in our study, on average, learned 50%–60% of the word–object pairs, a range similar to the findings in other studies. A recent cross-situational word learning study (Romberg & Yu, 2014) included a condition that had approximately the same amount of training trials as our study (58 vs. 54 in this study). Romberg and Yu (2014) found that learners' word learning performance was, on average, at ceiling (i.e., over 88% correct). These results suggest that word learning performance in our study overall was not as robust as studies containing only one level of co-occurrence regularities.

One possible cause of interference is that the words and objects used in this study were relatively similar to each other (i.e., stimuli from the same category shared certain perceptual features). Auditory or visual features shared among training stimuli might have caused difficulty in word or object recognition and discrimination and thus hindered word–object mapping (e.g., Edelman, 1995; Gauthier, James, Curby, & Tarr, 2003; Magnuson, Dixon, Tanenhaus, & Aslin, 2007). A second possible cause, somewhat related to

the first, comes from the effect of object categorization. It has been suggested that forming object categories can increase the perceived similarity among members within the same category and thereby increase difficulty in object discrimination (Goldstone, Lippa, & Shiffrin, 2001). This may be one of the reasons why Same-Category distractors affected Mapping performance. This may also explain why some participants had low Mapping but high Generalization performance. A third possibility is that participants may have noticed that there were two streams of regularities in the training stimuli, without necessarily mastering one or both of the regularities, and that these two levels of regularities competed for attention and processing. All three possibilities are plausible and not mutually exclusive. Some participants may experience one type of interference while others may experience two or even all three sources of interferences. Whatever the cause(s) may be, the distributions seen in Figs. 3 and 5 suggest that some learners were able to overcome this interference.

4.2. Individual differences in simultaneous word learning and category learning

This study revealed marked individual performance differences in word and category learning (Figs. 3 and 5). Participants' Mapping performance ranged from slightly below chance level to having perfect scores, with the majority of individuals having more correct Mapping trials than expected by chance. This indicates that most learners were able to perform the word learning task they were instructed to do. Learners with low Mapping scores tended to have low Generalization performance. Their low learning performance may be caused by difficulty in tracking or accumulating co-occurrence information at both word and category levels. This difficulty may be either caused by perceptual similarity among stimuli and/or due to competition and interference between two simultaneously presented statistical regularities. In contrast to overall fair to good Mapping performance, many participants' Generalization scores were on the lower end, suggesting that they may have either intentionally ignored the syllable-to-category associations present in the stimuli or simply failed to pick up or notice the presence of that stream of regularities. Although many participants only succeeded in the Mapping task, there was still a significant group of learners who performed above chance in both Mapping and Generalization tasks. Some of them learned a fair amount of word–object pairings and acquired some knowledge of syllable-to-category associations, whereas others performed at ceiling in both tasks. These learners, especially those performing at ceiling, seemed to overcome the interference and competition between different streams of regularities and may have used the information acquired from one type of regularity to help learn the other type of statistics in bootstrapping fashion.

Successful Generalization performance reflected sensitivity to the syllable-to-part associations. However, it is less clear whether learners who failed to generalize simply did not notice the category structures or whether their learning of syllable-to-category associations was not good enough to support generalization. One future direction of study to address this question would be to include eye-tracking measures and examine the gaze patterns during learning. The prediction is that participants who notice the category

structures in the stimuli will pay more attention to the critical object parts than those who do not. Tracking participants' gaze patterns may also reveal whether category learning bootstraps word learning. For example, learners who notice the syllable-to-category associations may use this cue to rule out irrelevant distractors and shift their gaze to the appropriate objects after hearing the critical category-relevant syllable.

One question raised by these findings is what caused individual performance differences? Previous cross-situational learning studies suggest that word learning performance can be affected by multiple factors, including learners' (selective) attention to different stimuli during learning (Smith & Yu, 2013; Yu, Zhong, & Fricker, 2012), their ability to aggregate and/or integrate information across trials (Romberg & Yu, 2013; Yu et al., 2012), and their memory capacities (Vlach & Johnson, 2013). Past research has shown that these abilities are crucial in category learning as well (e.g., Ashby & Gott, 1988; Blair, Watson, Walshe, & Maj, 2009; Lewandowsky, Yang, Newell, & Kalish, 2012). Furthermore, as learners were exposed to two concurrent streams of regularities in this study, it is likely that whether and how they selectively attended to these two separate streams of information during training and whether they could integrate information *across* streams affected word and category learning performance as well. These factors (separately or in combination) may have resulted in the individual performance differences observed in this study.

It should be noted that participants in this study were only instructed to learn word-to-object mappings, but not informed of the category structures. This design very likely directed participants' attention to the word learning aspect, and potentially away from the category structures present in the stimuli, as previous studies have shown that task instruction can drive participants' attention away from "irrelevant" information (e.g., Haider & Frensch, 1999). This may be one reason why learners generally had fair to good Mapping performance, but many of them did not perform well in the Generalization task. Changing the instruction in future studies to inform participants of the category structure or to tell them explicitly to pay attention to the shared features among individual items may improve category learning and facilitate generalization.

4.3. Initial syllable learning advantage?

The results of this study show that different training conditions did not significantly affect participants' Mapping and Generalization at the group level. However, the numerical patterns and chance analyses indicate that given the same co-occurrence information, English speakers had more robust category learning when the category markers were placed at the beginning of a word than at the end. A similar, albeit weaker, pattern was observed in the Mandarin speakers' performance as well. This finding is inconsistent with another recent study suggesting a final syllable advantage in learning grammatical categories (St. Clair, Monaghan, & Ramscar, 2009). When adult English speakers were presented with auditory stimuli containing regularities simulating either prefixing (i.e., prefix + root word, such as *ve-tweand*, *ve-dreng*) or suffixing (i.e., root word + suffix, such as *tweand-ve*, *dreng-ve*) structures, it was found that learners were able to group

together root words based on the associating suffixes, but not based on the prefixes. This result indicates a learning advantage when words belonging to the same grammatical category have the same final syllables.

In another study examining grammatical category learning, Monaghan and Mattock (2012) reported that adults were better able to learn word–object mappings when a word was preceded by a reliable category marker indicating whether the word was a referring or non-referring noun than when the marker was unreliable or not present. However, this study did not test whether placing the category markers after the target word would yield a similar effect. Although it is difficult to compare our results to these other studies because of experimental differences, together, the results suggest that in certain learning contexts there may be an initial syllable advantage in English (and possibly Mandarin) speakers' category learning while in other contexts, the learning advantage tends to fall on the final syllable. As mentioned previously, the informativeness or psychological saliency of the beginnings of English words likely draw English speakers' attention more toward the initial syllables during word processing and production (e.g., Cutler & Carter, 1987; Kessler & Treiman, 1997). It is possible that the same factors also contribute to the ordinally initial-syllable advantage shown in English speakers' (as well as Mandarin speakers') category learning. Future research is needed to examine how and when word structures affect category learning.

4.4. Future directions

One important direction for future research is the cross-linguistic aspects in statistical learning. The initial goal of our study was to investigate adults' general ability to extract two concurrent levels of statistical regularities by using linguistic stimuli analogous to structures that can be found in natural languages. Even though we recruited speakers of English and Mandarin, two languages which show very different prevalence levels of compound nouns, we recognize that, strictly speaking, these two languages may not provide a "clean" comparison. For example, although compounds are not as prevalent in English as in Mandarin, they are present in both languages; and the final syllables of English words can sometimes be predicative of category membership (e.g., *cheesecake* and *cupcake*). This may be one reason why we did not find significant group mean differences, but rather only minor pattern differences, in English and Mandarin speakers' Generalization performance. The data suggest that our findings about adults' overall word and category learning capability are fairly reliable. However, this work also indicates the necessity of future research to take participants' prior language or learning experiences into account.

Finally, it is worth noting that all participants in our study have some knowledge of at least a second language because second language courses are required in local high schools for our English-speaking participants and English courses are required for the Mandarin-speaking participants. Even though none of the participants was fluent in both English and Mandarin and they were trained and tested in their native language, we could not rule out the possibility that exposure to a second language (or second languages) may

have affected their learning of novel stimuli. Previous word learning studies using various experimental paradigms have shown differences in bilingual and monolingual individuals' attentional control and their mutual exclusivity bias (Byers-Heinlein & Werker, 2013; Poepsel & Weiss, 2016; Yoshida, Tran, Benitez, & Kuwabara, 2011). These factors likely also play a role in cross-situational word learning (e.g., Yurovsky et al., 2013). Another topic for future research is to further investigate how bilingualism or multilingualism affects attentional allocation and use of different word learning strategies or biases in statistical word learning.

5. Conclusions

This research demonstrates that adult learners are able to extract multiple levels of statistical regularities and use them to learn word-to-object mappings and to form object categories. Our work adds to a rapidly growing body of evidence that humans are skilled at accumulating and integrating information gathered across multiple modalities and across contexts and tasks. The overall comparable word and category learning performance from speakers of English and Mandarin, two typologically different languages, confirms the reliability of our findings. In addition, participants' learning patterns suggest that statistical word and category learning processes interact with each other.

Acknowledgments

Notes

1. The numerical number at the end of each syllable of the Mandarin stimuli indicates tonal information. The numerical number 1 indicates a high level tone; 2 refers to a rising tone; 3 is a falling-rising tone; and 4 indicates a falling tone.
2. It is noteworthy that even though final syllables are fairly reliable cues to category membership in Mandarin when compared to other languages (e.g., English), they are not deterministic cues. Objects from different taxonomic categories sometimes have homophonic final syllables. For example, the words *fei1-ji1* (airplane) and *mu3-ji1* (hen) have the same sounding final syllables, despite the fact that the written characters and meanings of these two head nouns *ji1* are different. Using the

final syllables of these two words as the *only cue* to judge the category membership of these two objects would result in miscategorization.

3. Our discussion about Mandarin adjectives is restricted to Attributive Adjectives and does not apply to Predicate Adjectives. Mandarin Predicate Adjectives are placed after the nominal headwords or after the Subject in a sentence and are viewed as a type of verb by some accounts (e.g., Chao, 1968; Cheung, Liu, & Shih, 1994; Li & Thompson, 1981). Here, we only consider the position of Mandarin Attributive Adjectives, which are placed before the head noun of a noun phrase.

4. According to GEPT-level descriptions, the first stage of the High Intermediate level test is to determine whether a participant is able to "understand English conversations in social settings and workplaces" and to "read different types of articles on concrete and abstract topics" (https://www.lttc.ntu.edu.tw/E_LTTC/E_GEPT/hi_intermediate.htm).

References

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33–53.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(02), 395–418.

Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *35*, 1196–1206.

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

Chao, Y. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Cheung, H. S., Liu, S., & Shih, L. (1994). *A practical Chinese grammar*. Hong Kong: Chinese University Press.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, *112*(2), 347–382.

Cunillera, T., Laine, M., Càmara, E., & Rodríguez-Fornells, A. (2010). Bridging the gap between speech segmentation and word-to-world mappings: evidence from an audiovisual statistical learning task. *Journal of Memory and Language*, *63*(3), 295–305.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3), 133–142.

Edelman, S. (1995). Class similarity and viewpoint invariance in the recognition of 3D objects. *Biological Cybernetics*, *72*(3), 207–220.

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

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

Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, *20*(3–6), 507–523.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Goldstone, R., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*, 27–43.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.

Haider, H., & Frensch, P. A. (1999). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied*, *5*(2), 129–151.

Huang, S. (1998). Chinese as a headless language in compounding morphology. In J. L. Packard (Ed.), *New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese* (pp. 261–283). Berlin, Berlin and New York: Mouton de Gruyter.

Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes*, *24*, 876–909.

Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies: Determining the speech of lexical access for Mandarin and English compounds. *Language and Cognitive Processes*, *23*, 1191–1223.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2013). Actively learning object names across ambiguous situations. *Topics in Cognitive Science*, *5*, 200–213.

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language*, *37*(3), 295–311.

Lewandowsky, S., Yang, L. X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 881.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. D., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*, 1–24.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77.

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

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

Monaghan, P., & Mattock, K. (2012). Integrating constraints for learning word–referent mappings. *Cognition*, *123*(1), 133–143.

Plag, I., Kunter, G., & Lappe, S. (2007). Testing hypotheses about compound stress assignment in English: A corpus-based investigation. *Corpus Linguistics & Linguistic Theory*, *3*(2), 199–232.

Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*, 665–681.

Poepsel, T. J., & Weiss, D. J. (2016). The influence of bilingualism on statistical word learning. *Cognition*, *152*, 9–19.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science*, *37*, 1290–1318.

Romberg, A. R., & Yu, C. (2013). Integration and inference: Cross-situational word learning involves more than simple co-occurrences. In M., Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Romberg, A. R., & Yu, C. (2014). Interactions between statistical aggregation and hypothesis testing mechanisms during word learning. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6038–6043.

Sloutsky, V. M., Lo, Y., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development*, *72*, 1695–1709.

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Sciences*, *35*, 480–498.

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

Smith, L. B., & Yu, C. (2013). Visual attention is not enough. *Language, Learning, and Development*, *9*(1), 25–49.

St. Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science*, *33*(7), 1317–1329.

Suanda, S. H., & Namy, L. L. (2012). Detailed behavioral analysis as a window into cross-situational word learning. *Cognitive Science*, *36*, 545–559.

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science*, *34*, 1093–1106.

Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, *10*(2), 201–224.

Tsai, J.-L., Lee, C. Y., Lin, Y.-C., Tzeng, O. J. L., & Hung, D. L. (2006). Neighborhood size effects of Chinese words in lexical decision and reading. *Language and Linguistics*, *7*, 659–675.

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

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*(3), 217–242.

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112*, 97–104.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245–272.

Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual children. *Frontiers in Psychology*, *2*, 210. doi:10.3389/fpsyg.2011.00210.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Review*, *119*(1), 21–39.

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective attention in cross-situation statistical learning: Evidence from eye tracking. *Frontiers in Psychology*, *3*, 148. doi:10.3389/fpsyg.2012.00148.

Yurovsky, D., Yu, C., & Smith, L. B. (2012). Statistical speech segmentation and word learning in parallel: Scaffolding from child-directed speech. *Frontiers in Psychology*, *3*, 374. doi:10.3389/fpsyg.2012.00374.

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, *37*, 891–921.

Zhou, X., Marslen-Wilson, W., Taft, M., & Shu, H. (1999). Morphology, orthography, and phonology in reading Chinese compound words. *Language and Cognitive Processes*, *14*(5/6), 525–565.