What’s in a Name, and When Can a [Beep] be the Same?

What’s in a Name, and When Can a [Beep] be the Same?

Jill Lany1,2, Abbie Thompson3, & Ariel Aguero1

1 University of Notre Dame, Department of Psychology

2 University of Liverpool, Department of Psychological Sciences

3 Valparaiso University, Department of Psychological Sciences

This research was funded by NSF Grant BCS 1352443 to Jill Lany. We

thank the families who participated in this research, and members of the

Infant Studies Lab for their contributions to this work.

Correspondence concerning this article should be addressed to Jill Lany,

Department of Psychological Sciences, University of Liverpool, Bedford

Street South, Liverpool L69 7ZA, United Kingdom. Email: jlany@liverpool

.ac.uk

**Abstract**

Words influence cognition well before infants know their meanings. For example, three-month-olds are more likely to form visually-based categories when exemplars are paired with spoken words than with sine-wave tones, a likely precursor to learning symbolic relations between words and their referents. However, it is unclear why words have these effects. In three experiments we tested the hypothesis that exaggerated “showing” gestures used when naming objects, and the resultant cross-modal synchrony between a name and object motion, can affect object categorization. Participants were 119 3-month-old infants (56 were female and 63 were male). According to caregiver report, the sample was composed of European American (N = 114) Black (N = 6), Hispanic (N = 2) and multi-racial (N = 6) infants. Participants were growing up predominantly in homes with at least one parent who completed a college education or a higher degree (80%), and the remaining 20% completed high school. After replicating evidence that words and tones have different effects on categorization, we found that pre-familiarizing infants with tone-object synchrony leads tones to influence categorization as words do. Moreover, we found that concentrated experience with word-object synchrony enhances the effects that words themselves have on categorization. Thus, temporal structure within caregivers’ communicative behaviors may lead words to facilitate categorization, and ultimately to aid in forming symbolic representations.

**Keywords:** categorization, amodal relations, language and thought, language acquisition

For mature speakers of a language, words are not merely associated with referents, but rather serve as symbols for them, raising the question of how infants come to form symbolic relations over associations when learning language. Remarkably, words bear a special relation to concepts from very early in development: Infants who are just a few months old are more likely to form visually-based categories when exemplars are paired with naturally-produced human speech than when they are paired with most other auditory stimuli, such as sine-wave tones or backwards speech (Ferry et al., 2010; 2013). For example, when 3-month-olds viewed a series of dinosaur exemplars, each one paired with the same word, in the subsequent test they preferred a new dinosaur exemplar to a fish when they were presented side-by-side, even though both images were novel to them. This preference is referred to as a familiarity preference, in that it is a preference for an exemplar from the familiarized category, and is taken evidence that infants formed a dinosaur category. Critically, when 3-month-olds viewed the same series of dinosaur exemplars, each paired with sine-wave tones, they subsequently failed to show a preference for either test exemplar, suggesting that words facilitate category formation relative to tones. Initially the link between words and categories is relatively general, such that any word promotes object-based categorization, regardless of its form (Waxman & Booth, 2003). Eventually the link is differentiated and refined, such that different kinds of words are more selectively linked to different kinds of concepts (i.e., nouns highlight object kinds in categorization tasks, while adjectives highlight object properties; Booth & Waxman, 2001).

These results suggest that words play a formative role in object cognition, and that an early link between words and categories sets the stage for learning symbolic relations. But where does this link come from? One possibility is that it is experience-independent. At least some nonhuman primate vocalizations, which are unfamiliar to human infants, have the same beneficial effects on categorization that words have (Ferry et al., 2013). Thus, the neural circuits that are engaged by the acoustic properties of speech and other similar signals could be prewired to activate the systems involved in category-formation and representation.

However, we hypothesize that infants’ experience with the structured relations between words and objects that are generated in communicative interactions can lead words to facilitate categorization. Human communication often contains striking correspondences between information streams that are conveyed across different modalities (McNeill, 1992). For example, words and gestures that convey related information are often produced on the same beat (i.e., in synchrony), potentially reflecting entrainment in the motor system (Kelso et al., 1983), or shared dependence on a self-organizing process driving communication (McNeill, 1992). This kind of coordination across modalities is exaggerated in speech to infants, with caregivers frequently presenting objects and their labels in synchrony, shaking an object or looming it towards an infant just as they utter the word for it (Gogate et al., 1998; Iverson et al., 2006).

Critically, such cross-modal synchrony promotes the unification of distinct sensory inputs that are generated by the same underlying event (Bahrick & Lickliter, 2014). For example, infants are more likely to encode a rhythmic pattern within an event like a hammer tapping on a surface when it is specified multimodally (i.e., the timing regularities that are present in both visual and auditory modalities) vs. unimodally (Bahrick et al., 2002). Unlike the tapping event, synchrony between spoken words and object motion in caregiver communication does not reflect an inherent connection or unity. Instead, caregivers spontaneously create this structure as they show and talk about objects with their infants. Nonetheless, encoding synchronously occurring words and objects may recruit amodal processing mechanisms whose predominant function is imparting tight connectedness, and even equivalence, on information from different sensory modalities. Thus, at a macroscopic level, caregivers’ behaviors could yield a precursor to symbolic representations by imposing unity or equivalence on infants’ representations of words and objects.

At a more mechanistic level, synchrony could promote categorization by changing how infants encode visual information. Heightened attention to redundant information in multimodal events often comes at the cost of attention to information that is present in just one sensory modality, especially early in development (Bahrick & Lickliter, 2014). For example, in the hammer-tapping event described above, while infants benefit from amodal information when learning rhythmic structure, they are better able to encode information about the hammer that is only specified visually (i.e., its orientation) when it is experienced unimodally (Bahrick et al., 2006). Thus, the presence of word-object synchrony may lead to diminished visual encoding, or to visual representations of objects that are heavily dominated by properties that they encode most readily early in development, such as shape (Wilcox, 1999), an excellent cue to category membership (Biederman, 1987).

Alternatively, word-object synchrony could lead to the formation of stronger or more veridical object representations. Although attention to synchrony can diminish unimodal encoding early in development, as infants get older, and/or better able to encode synchrony, they also become better able to encode the modality specific properties of such events. For example, at 4 weeks of age infants rely on audio-visual synchrony to identify which of several objects made a sound (i.e., synchrony between sound onset and an object contacting a surface or another object, being squeezed, or shaken, etc.; Bahrick, 2001). Critically, attention to this kind of synchrony can support learning about more specific properties of such events by 7 weeks of age, such as what kinds of sounds the objects make (i.e., whether they make a unitary or more complex sound when contacting a surface). By 7 months of age, infants can use synchrony to learn even more arbitrary relations, such as the pitch an object of a particular color makes upon impacting a surface (Bahrick, 1994). ERP data also suggests that the presence of audio-visual synchrony in videos of talking faces promotes accurate encoding (Reynolds et al., 2014). Thus, there are at least two ways in which synchrony could shape what is learned about objects encountered during naming, and in turn shape whether and how they are grouped into categories.

Note that studies showing that words have a special effect on categorization have not generally involved word-object synchrony. Instead, words had these effects all on their own. Critically, exposure to amodal structure in multimodal events can actually scaffold attention to relevant information in subsequent events lacking synchrony. A powerful example comes from a study of bob-white quail embryos. When exposed to maternal calls in synchrony with flashing lights, the quail were better able to encode their temporal structure, and crucially, to perceive this same temporal structure in a subsequent unimodal event (Lickliter et al., 2006). This effect persisted across delays of up to 4 hours, suggesting that experience with amodal processing can have lasting effects on attention and learning. Similar developmental scaffolding appears to occur with human infants. For example, experience with affect expressed simultaneously in faces and voices may facilitate sensitivity to affect when it is expressed unimodally (Flom & Bahrick, 2007; Walker-Andrews, 1997). Furthermore, infants generalize expectations about synchrony to new objects and events by 3.5 months (Bahrick, 2002). Words could likewise come to influence visual encoding by virtue of prior synchrony with object motion. Thus, the effects of words on categorization documented by Ferry et al. (2010) and others, which did not involve word-object synchrony, could have been shaped by a history of it.

Given these findings, we hypothesized that sound-object synchrony can lead that sound type to facilitate object categorization. In Experiment 1 we first attempted to replicate Ferry et al. (2010), who found that words have special effects on categorization, relative to tones, in 3-month-old infants. Specifically, as described above, pairing category exemplars with words during a familiarization phase led infants to prefer a novel exemplar from the familiarized category over an exemplar from an unfamiliar category at test, but when the exemplars were paired with sine-wave tones, they showed no preference.

Building on the replication in Experiment 1, in Experiment 2 we asked whether synchrony between a sound and object motion can impact how that sound subsequently affects 3-month-olds’ categorization. To do so, we tested whether giving infants prior experience in which tones occur in synchrony with object motion leads tones to behave like words in a subsequent categorization task. Specifically, we prefamiliarized infants with instances of tone-object synchrony, next familiarized them with category exemplars paired with tones, and finally tested whether they preferred a novel exemplar from the familiarized category over an exemplar from an unfamiliar category in a subsequent test phase.

In Experiment 3 we tested whether experience with word-object synchrony influences the effects of words on object categorization at 3 months. We based our experimental logic on evidence that words continue to boost categorization as infants get older, but the way that successful categorization is manifested changes. Specifically, 4-month-old infants familiarized with category exemplars paired with words showed a preference for the exemplar from the *novel category*, referred to here as a novelty preference (Ferry et al., 2010), while 3-month-olds given the same categorization task preferred the exemplar from the *familiarized category*. Importantly, both familiarity and novelty preferences can indicate learning, but novelty preferences are more likely to emerge in older infants and in easier tasks, such as those involving more familiarization (Hunter & Ames, 1988), or more retrieval cues (Bahrick et al., 1997). Thus, the switch from a familiarity preference at 3 months to a novelty preference at 4 months observed by Ferry et al. (2010) suggests that the categorization task is relatively challenging for 3-month-olds, even with words, and that the effect of words on categorization actually gets stronger with time and experience. Thus, we tested whether giving 3-month-olds intensive experience with word-object synchrony leads them to show a novelty preference in a subsequent categorization task in which exemplars are paired with words.

**Experiment 1**

Ferry et al. (2010) found that 3-month-olds perform differently on an object categorization task when exemplars are paired with words vs. sine-wave tones during the training phase. In Experiment 1, a replication of that work, we likewise tested whether 3-month-olds familiarized to exemplars paired with words subsequently showed a greater familiarity preference than those familiarized to the same exemplars paired with tones. Critically, this replication also provides key comparisons for the data from Experiments 2 and 3, in which infants were pretrained with synchronous tones and words before being given the categorization task.

**Methods**

**Participants.** Our participants were 38 monolingual, English-learning infants who were born full-term, had normal hearing and no history of chronic ear infections or serious health or cognitive issues. Of the 38 infants, 2 were Black, 1 Hispanic, 2 mixed race, and the remaining 31 were European American, according to parent reports. Seventeen of the infants were female, and the average age was 3;13 months (the range was 3;0 months to 3;27 months). This study (and all experiments within it) was approved by the Institutional Review Board of the University of Notre Dame under the name Learning Mechanisms in Infant and Child Language Development, and the number 17-11-4219.

We used G\*Power 3.1 (Faul et al., 2009) to compute our goal sample-size using an *a priori* analysis, planning for a one-sample *t* test comparing two groups. and 85% power. We used the effect size reported from Ferry et al.’s (2010) work with infants of the same age (a Cohen’s *d* of .87), and which used a highly similar categorization paradigm to compare performance in Word and Tone conditions. This yielded a goal sample of 20 infants per condition. We tested several more infants than our goal sample in both conditions to accommodate for the need to exclude those who were inattentive during the Categorization Training phase (1 infant was excluded in the Word condition, and 5 in the Tone condition, for this reason). Data from additional infants were excluded because of fussiness (N = 5), equipment failure (N = 1), or experimenter error (N = 1). Our final sample sizes were 22 in the Word condition, and 16 in the Tone Condition. Note that the Tone group was smaller because more infants had to be excluded for inattention.

**Materials.** All infants were given a categorization task consisting of a Training and Test phase, following the design of Ferry et al. (2010). We modeled our materials after those used by Ferry et al. as closely as possible, but did not use their exact auditory and visual stimuli, instead creating our own versions. The visual materials for the Training phase consisted of line-drawn pictures of exemplars from two different animal categories (dinosaurs and fish; see Figure 1). Within a category there were 8 training exemplars, each of which was outlined in black and filled with a unique color. The same colors were used in each of the two categories (i.e., there was a lime-green dinosaur and a lime-green fish). Half of the infants in each condition were trained on dinosaur exemplars, and half on fish exemplars.

The auditory materials used in the Training phase were two sentences containing the nonsense word “modi” or “toma” (e.g., “Look at the modi.” “Do you see the modi?” or “Look at the toma.” “Do you see the toma?”) in the Word condition, and sine-wave-tone-sequences of either 400 or 800 Hz in the Tone condition. The tone sequences were matched in duration and amplitude to the speech sequences. We counterbalanced which of the two words or tone frequencies infants heard.

The Test trial materials consisted of two exemplars presented side-by-side: a within-category exemplar (e.g., a drawing of a dinosaur that had not been presented during Training) was presented next to an image of the same color from a novel category in silence (e.g., a fish; see Figure 1). Thus, the dinosaur was the familiar category exemplar for half the infants, and the novel category exemplar for the other half. There were two sets of test images (i.e., a powder-blue fish and dinosaur, and a yellow fish and dinosaur, both colors that had not been used during the Training phase). In one set the fish was positioned on the right, and in the other set it was on the left. As described below, infants were given two Test trials and thus saw both image sets. The order in which the individual Test trials were presented was counterbalanced across infants.

Figure 1

*Design of the Categorization Training and Test Phases*



*Note.* Figure created by authors.

**Procedure.** Infants were randomly assigned to the Word or Tone condition. In both conditions, infants participated in the categorization task modeled after Ferry et al. (2010), seated on a caregiver’s lap approximately 1 meter from a 60-inch LCD monitor, with speakers and a camera mounted underneath. The categorization task consisted of a Training phase and a Test phase, the latter using a visual-paired comparison (VPC) procedure. In both conditions, the Training phase consisted of a series of 8 exemplars from one of the animal categories displayed on the screen one at a time, with each image visible for 20 seconds. Each exemplar appeared in silence for 8 seconds, and then an auditory stimulus (a word or tone sequence described above, depending on the condition) was presented. After another 8 seconds the auditory sequence was repeated. In the Word condition, exemplars were paired with one of the words for a given infant, and in the Tone condition they were paired with one of the tone frequencies. Thus, for an infant in the Word condition each exemplar was paired with a word (e.g.) “modi”, and for an infant in the Tone condition each exemplar was paired with beeps of the same sine-wave tone frequency (see also Figure 1).

In the Test phase, each trial began with a video of a flashing light, which was presented until the experimenter was certain that the infant was attending to the screen, and the test stimulus (a novel exemplar from the familiar category side-by-side with a novel exemplar from the unfamiliarized category) was presented in silence until infants accumulated 10 seconds of looking. Infants were given two test trials, and the image set that was presented first was counterbalanced across participants (i.e., for half of the infants the powder-blue fish and dinosaur set were seen first, and for the other half the yellow set was seen first). Note that the test trials were identical for infants in the Word and Tone Conditions; the only thing that differed across the conditions was whether the training exemplars were paired with words or tones.

Altogether there were multiple factors that were counterbalanced across infants; the training set (i.e., whether they were trained on dinosaurs or fish), which of the two tones or two words was paired with the images during Training, the test images presented in the first vs. second Test trial (i.e., the yellow or blue set), and the side of the screen on which the target image was presented first (right or left) at Test.

Infants were video recorded for the duration of the experiment, and their looking behavior was coded frame-by frame offline using iCoder (Fernald et al., 2008). During the Training phase coders indicated, for each frame, whether the infant was looking to the display, or looking away, to get a measure of attentiveness. During the Test, coders indicated whether the infant was looking to the image on the left, the image on the right, transitioning between pictures, or looking away from the display. Looking behavior on only the first test trial was analyzed, except if the infant looked exclusively at one image the entire 10 seconds (i.e., looking to one of the objects 100% of the time, never looking to the other), in which case the second trial was used, following the approach of Ferry et al. (2010; 2013). The data from of 15% of infants was recoded, and agreement was 98% across coders. Infants who were inattentive for 80% or more of the Training phase were excluded from analysis. This study was not preregistered, and the stimuli and data are available upon request.

**Results**

Our primary question was whether infants’ performance on the categorization Test differed depending on whether category exemplars were presented with words or tones during Training. Following Ferry et al. (2010), performance on the Test was quantified as the proportion of time that infants spent looking to the exemplar from the novel category divided by the sum of looking to both the novel and familiar category exemplars. Thus, scores above .5 on this measure reflect preferential attention to the exemplar from the novel category (a novelty preference) and those below .5 reflect preferential attention to the exemplar from the familiarized category (a familiarity preference). Note that because there were two exemplars presented side-by-side at test, this single score reflects relative attention levels to both the exemplar from the familiar category and the exemplar from the novel category. We have used this preference score metric (referred to as a Novelty Preference Score in the figures and tables to indicate that larger values on this metric reflect a larger novelty preference) to be consistent with Ferry et al. (2010). However, note that we could just as easily have reported this preference score such that larger numbers reflected a familiarity preference (and this score can easily be computed as 1- the novelty preference score).

In this paradigm, preferential attention to the novel exemplar from the familiarized category, which is what we predicted based on the results of Ferry et al. (2010), is likely to reflect infants’ ability to detect the similarity of that exemplar to the familiarized ones. In particular, given that both of the test exemplars, presented side by side, are novel (in terms of their shape, features, and color) infants could not show a familiarity preference on the basis of an exact memory match. Instead, the greater similarity of the test exemplar from the familiar category, in terms of shape and other relevant features would attract their attention. However, memory is clearly a factor, in that infants would have to remember something about the exemplars across each familiarization trial to perceive category-based similarities across them, and their memory representation would have to be flexible enough to be activated by nonidentical and unfamiliar category exemplar.

We used both traditional *t* tests and Bayesian *t* tests to determine whether the Word and Tone conditions differed. Both types of analyses were performed using JASP (JASP Team, 2019), with prior distribution set to the Caushy default of .707. The Bayesian *t* tests yield Bayes Factor (BF) values that reflect the strength of evidence in favor of both the alternative and null hypotheses (BFalt or BFnull)), given the observed data. BF values above 1 are interpreted as anecdotal evidence in support of a given hypothesis, and scores above 3 can be interpreted as moderate evidence in its support (Lee & Wagenmakers, 2014). The results of the Bayesian analyses, along with effect sizes, can therefore supplement the *t* test results, and are especially relevant for interpreting when a lack of difference is reliable (i.e., a relatively large BFnull), offering an alternative to measures of observed power. We also report observed power, which was computed using *post hoc* analysis in G\*Power (Faul et al., 2009), with our observed effect sizes (Cohen’s *d*), and sample sizes.

Ferry et al. (2010) found that 3-month-olds show a greater preference for the within-category exemplar (i.e., a familiarity preference, or scores less than .5) in the Word Condition than in the Tone condition. Thus, we predicted that we would also find a greater familiarity preference in the Word condition. *Numerically* infants in the Word condition showed a greater familiarity preference than infants in the Tone condition (see Table 1 for mean preference scores in both of these conditions). The groups did not significantly differ when they were directly compared using a directional one-sample *t* test; *t* (36) = 1.54, *p* = .066, *d* = .381, observed power = .306). However, the effect size was suggestive of a group difference. A Bayesian *t* test yielded anecdotal support for the hypothesis that infants in the Word condition showed a greater familiarity preference than those in the Tone condition (BFalt = 1.451), and little support for the null hypothesis that they did not differ (BFnull = .689). Thus, there was only weak evidence that the two groups differed.

Table 1

*Novelty Preference Scores on the Categorization Task*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 10 Seconds | | |  | First 5 Seconds | | |  | Second 5 Seconds | | |
|  | *M* | *SE* | *d* |  | *M* | *SE* | *d* |  | *M* | *SE* | *d* |
| Word | 0.430 | 0.061 | 0.247 |  | 0.427 | 0.066 | 0.236 |  | 0.445 | 0.081 | 0.218 |
| Tone | 0.569 | 0.065 | 0.269 |  | 0.637 | 0.082 | 0.418 |  | 0.484 | 0.077 | 0.052 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Async Tone | 0.532 | 0.058 | 0.120 |  | 0.527 | 0.067 | 0.091 |  | 0.546 | 0.072 | 0.140 |
| Sync Tone | 0.422 | 0.055 | 0.320 |  | 0.401 | 0.062 | 0.357 |  | 0.420 | 0.074 | 0.247 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Sync Word | 0.592 | 0.052 | 0.389 |  | 0.607 | 0.076 | 0.310 |  | 0.572 | 0.070 | 0.224 |
| Async Word | 0.471 | 0.067 | 0.100 |  | 0.385 | 0.082 | 0.322 |  | 0.439 | 0.090 | 0.156 |

*Note*. Scores reflect the mean proportion of time infants spent looking to the novel exemplar from the unfamiliar category at test. The effect sizes reflect the extent to which performance differed from chance (or .5).

The failure to find clear evidence that words and tones have different effects on categorization was unexpected, as many studies have reported such differences with large effect sizes, even at 3 months of age (e.g., Ferry et al., 2010). However, we did find modest evidence of a difference in performance that is consistent with that found in prior research, in terms of the Bayes Factor and the effect size. Given that infants can recognize shared category membership within 2-3 seconds of viewing a stimulus (Althaus & Mareschal, 2014), and that these effects can be short-lived (LaTourrette & Waxman, 2018), we tested whether the Word condition showed a greater familiarity preference in the first 5 seconds of the 10-second test trial in an exploratory analysis. This analysis revealed that infants in the Word condition had a stronger familiarity preference than those in the Tone condition (directional *t* (36) = 2.015, *p* = .026, *d* = .659, observed power = .627). The effect size was nearly double that of the full 10 seconds, and a Bayesian *t* test provided stronger evidence in support of the hypothesis that the conditions differed in the predicted direction (BFalt = 2.876 and BFnull = .348). However, in the second 5 seconds there was no such evidence (directional *t* (36) = .369, *p* = .74, *d* = .113, observed power = .096; BFalt =.41 and BFnull = 2.44). The small effect size and the relatively large BF null both suggest that performance differences had faded by the second 5 seconds of the test trial. Thus, we were able to replicate the finding that words and tones have different effects on categorization at 3 months of age within the first 5 seconds of the test trial.

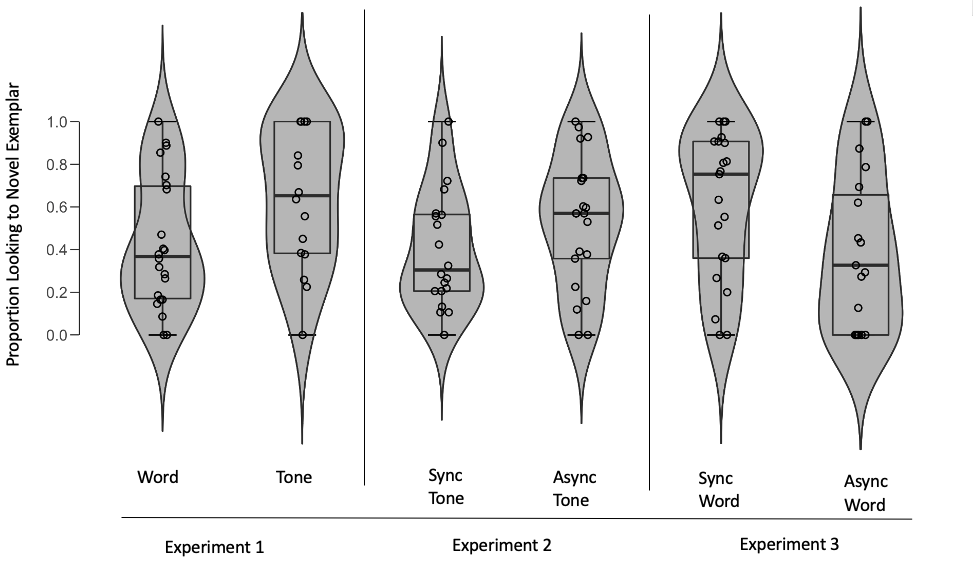
Ferry et al. (2010) found that infants in the Word group showed a significant familiarity preference at test (in other words, a mean score significantly below .5), indicating that they looked more to the exemplar from the familiar category than to the one from the unfamiliar category, but that infants in the Tone condition looked to the two exemplars equally. In the current experiment, although infants in the Word group showed a significantly stronger familiarity preference than infants in the Tone group, neither of the groups’ preference scores differed significantly from chance (or .5) during that time, (*t*s < .167, *p*s > .1; see Table 1 for means and effect sizes, and Figure 2 for violin plots). Ferry et al. also found that the number of infants in the Word condition showing a familiarity preference was greater than chance using a binomial test. We found that 15 of the 22 infants (68%) in the Word condition showed a familiarity preference, and that only 6 of 16 (38%) in the Tone condition did, but these frequencies did not differ from chance in a binomial test in either condition (*p*s = .134 and .454, respectively).

Thus, in contrast to the findings of Ferry et al. (2010), infants in the Word condition in the current experiment did not show a particularly strong familiarity preference, both when considering the results of the *t* tests comparing performance to chance, or the binomial tests comparing the number of infants showing a familiarity preference to chance. A similar pattern, in which the categorization performance of infants hearing words vs. tones differed, but the performance of those hearing words did not differ from chance, has also been reported in prior studies (Balaban & Waxman, 1997), and thus these results are not inconsistent with the existing evidence on how the different effects of words and tones manifest in a categorization task. However, in the current experiment, the effect size in the Tone group was larger than that in the Word group, suggesting they may have been doing something systematic rather than simply failing to categorize. We will consider this issue further in the General Discussion.

We next tested whether there were differences in the amount of attention infants in the two conditions payed during the Categorization Training phase. Infants in the Word condition attended to the display for 57.8% of the time, on average, and infants in the Tone condition attended for 48.0% of the time. A test of the hypothesis that infants in the Word condition were more attentive did not meet significance (directional *t* (36) = 1.672, *p* = .052, *d* = .549, observed power = .498), but the effect size is suggestive of a difference. Likewise, the Bayesian analyses provided anecdotal evidence that infants in the Word condition were more attentive (BFalt = 1.734 and BFnull = .577). Furthermore, more infants in the Tone condition than in the Word condition had to be excluded for inattention. Altogether, these results are consistent with evidence that infants often show heighted attention to objects in the presence of words (Baldwin & Markman, 1989). In fact, infants sometimes, though by no means always, show heightened attention to objects when they are paired with words than when they are paired with nonword stimuli in categorization tasks such as this one (e.g., Erickson, et al., 2014).

Figure 2

*Test Performance in Experiments 1-3*



*Note*. Violin plots depicting looking behavior in terms of Novelty Preference scores, or the proportion of time infants spent looking to the exemplar from the unfamiliar category. Scores above .5 indicate a preference for the exemplar from the novel category and scores below .5 indicate a preference for the exemplar from the familiar category.

Thus, it is reasonable to ask whether differences in attention to the training exemplars, rather than differences in the effects of tones and words on categorization, are centrally responsible for the observed differences in performance on the categorization task. For example, it is possible that greater attention to the familiarization exemplars, rather than the presence of words, led to a stronger familiarity preference in the Word condition at test. Note that it is possible that words have their effects by increasing attention to exemplars, but greater attention during categorization training in the Word vs. Tone conditions still makes it difficult to rule out the possibility that attention alone led to differences in test performance. However, it is unlikely that simple attentional differences alone can explain performance differences across conditions for a few reasons. First, there were no correlations between the amount of attention paid during training and test preferences, in either group alone (Word: *r* = .17, *p* = .444 and Tone: *r* = .297, *p* = .263, or in the two groups together (*r* = .119, *p* = .476). Furthermore, to preview the results of Experiments 2 and 3, there was no evidence of differences in attention during the categorization training phase in groups that differed in categorization performance, or of correlations between attention during training and test performance.

**Discussion**

In this experiment we replicated previous findings that words and tones have different effects on infants’ categorization performance by 3 months of age (Ferry et al., 2010). However, these effects appeared to be short-lived in comparison with those reported in previous studies, as we found that infants in the Word condition showed a greater familiarity preference than those in the Tone condition only within the first 5 seconds of the 10 second test trial, while previous studies have found a difference when considering the entire 10 second window. Furthermore, infants in the Word condition showed more modest evidence of categorization than those in previous studies, with small-to-medium-sized familiarity preference. Nonetheless, the differences we found are in generally line with previous results (Ferry et al., 2010). While we based our materials and procedure around that of Ferry et al., we did not use identical stimuli, and did not exactly match the size of presentation, or overall lab set up. All of these differences may have led to the slightly weaker effects we observed.

In two subsequent experiments we tested whether pretraining 3-month-olds with tone-object synchrony (Experiment 2) and word-object synchrony (Experiment 3) can change the effects that tones and words have on object categorization. This approach allowed us to test specific predictions about how performance should or should not change, relative to the original Tone and Word conditions, by manipulating whether the pretraining did or not did not involve synchrony. Furthermore, by limiting our analyses to the first 5 seconds of the test trial in our follow-up experiments, we were able to test whether evidence of categorization is consistently found in this time window (i.e., whether we could replicate the key effects found in the first 5 seconds in all subsequent tests).

**Experiment 2**

In Experiment 2 we tested the prediction that giving infants experience with tones that occur in synchrony with object motion can make tones behave like words in a subsequent categorization task. Specifically, we asked whether pretraining with tone-object synchrony can lead infants to show a greater familiarity preference on that task than infants in the original Tone condition. To that end, we created a Pretraining Phase in which infants viewed a video consisting of bouts of object motion accompanied by sine wave tones. In the Synchronous condition, tones occurred in synchrony with object motion. To address the possibility that any pretraining experience with tones could lead to changes in their impact on categorization, we also included an Asynchronous condition in which infants heard the same tones paired with the moving objects, but in which the temporal synchrony between them was disrupted. We used a mechanical claw to move the objects in the Pretraining phase in order to isolate temporal synchrony from other communicative features of interactions, such as the presence of a human, given evidence that infants do not perceive interactions between a mechanical claw and objects as having a psychological cause (e.g., Woodward, 1998).

**Methods**

**Participants.** Our participants were 41 3-month old infants, drawn from the same population as those tested in Experiment 1. Infants’ mean age was 3;15, the range was 3;1 to 3;30, and 20 were female. Of the 41 infants, 37 were European American, 2 Black, 1 Hispanic, and 1 mixed race, according to parent report. We used the same target sample sizes as in Experiment 1 so that the results across the experiments could be compared. Infants were assigned to the Synchronous (N = 20) or Asynchronous (N = 21) condition. Additional infants were excluded for fussiness (N = 3) and equipment failure (N = 1). We also excluded 1 infant (in the Asynchronous Tone condition) for inattention at test, and 2 infants (1 in the Asynchronous Tone and 1 in the Synchronous Tone condition) for inattention during the Pre-Training video.

**Materials.** The materials for the Pretraining phase were a series of 6 videos, each 20 seconds long. They depicted a mechanical claw made of a silver metal rod and black pinchers grasping and moving a set of objects. Each video began with one of the objects centered on a white stage-like background. The claw entered from the lower portion of the screen and moved up towards the object. The pinchers then grasped the object (see Figure 3 for a frame from one of the videos, and the full Synchronous and Asynchronous videos can be found in the Supplemental Materials) and moved it side-to-side in a shaking motion, or loomed it towards and away from the viewer. The objects were 3 chew-toys for dogs, chosen to be distinctive and relatively unfamiliar to infants. The movements, however, were modeled after object-demonstration behaviors that caregivers commonly employ during labeling (Gogate et al., 1998). After repeating the shaking or looming motion 4 more times, the claw released the object and moved back down and out of view. Then the next video began. Across the Pretraining phase, each of the three objects occurred once in a shaking video and once in a looming video, for a total of 6 videos. We created two sets of 6 videos that differed only in the duration of the intervals between the repetitions of the shaking and looming motions. Sine-wave tones of 293 Hz (a frequency not used in the subsequent Categorization Training phase) with a duration of approximately 1 second were overlaid onto the videos.

Figure 3

*Pretraining Phase Display*



*Note.* This image, created by the authors, displays a still shot from a Pretraining video containing the claw and one of the objects (see the Supplemental Materials for the full videos).

The critical manipulation was in how the tones were aligned with the object motion. In the Synchronous Tone condition, each sine-wave tone was timed to occur in synchrony with a bout of object motion. For example, a tone persisted for the duration of the shake, or was centered around the apex of the loom. In the Asynchronous Tone condition, the tone sequences that were synchronized to object motion intervals from one video set were overlain on the other set, such that the tones were not systematically aligned with the bouts of object motion. We counterbalanced which video infants saw during Pretraining.

**Procedure.** During Pretraining, infants were shown the Synchronous or Asynchronous videos, which took approximately 2 minutes. Infants were then given the categorization Training and Test, following the procedure in Experiment 1. In both conditions, tones were used as the auditory stimuli in the Training phase, which was identical to that used in the Tone condition from Experiment 1.

**Results**

The data were coded offline as in Experiment 1. Reliability across coders, calculated as in Experiment 1, was 96%. Based on the results from that experiment, we included only the first 5 seconds of looking behavior during the test trial in our analyses (see Table 1[[1]](#footnote-1)). This allowed us to determine whether we could replicate the finding that differences in categorization are observed within that timeframe. Figure 2 depicts categorization performance in the two conditions, alongside the Tone and Word conditions from Experiment 1.

We first asked whether infants in the Synchronous Tone condition showed a greater familiarity preference than those in the Tone condition from Experiment 1. A directional *t* test indicated that they did; *t* (34) = 2.335, *p* = .013, *d* = .775, observed power = .789. Furthermore, a Bayesian *t* test provided moderate evidence in support of the hypotheses that these groups differed in the predicted direction (the BFalt = 4.862, and the BFnull = .206). If the difference in categorization performance observed in the Tone and Synchronous Tone conditions reflects a general effect of giving infants pre-exposure to tones, then infants in the Asynchronous Tone condition should also show a greater familiarity preference than infants in the Tone condition. However, a directional *t* test indicated that this was not the case; *t* (35) = 1.04, *p* = .153, *d* = .345, observed power = .288. The Bayes Factors did not suggest any evidence in favor of the hypothesis that they differed (BFalt =.799) and provided anecdotal evidence that they did not (BFnull =1.251). These results suggest that only pretraining with synchronous tone-object motion led infants to perform differently from those with no prior training with tones.

Figure 2 shows that the performance of infants in the Synchronous Tone condition was very similar to that of infants in the Word condition from Experiment 1. Indeed, a directional *t* test failed to reveal a significant difference between the Synchronous Tone and Word conditions (*t* (40) = .281, *p* = .61, *d* = .087, observed power = .085), and a Bayesian *t* test provided moderately strong evidence that they did not differ (the BFalt =.25 and BFnull = 3.975). Altogether, these findings suggest that experience with tone-object synchrony changes how tones subsequently influence categorization, and leads them to have effects similar to those of words, while additional experience with tones without synchrony does not clearly do so.

Another way to test the hypothesis that experience with tone-object synchrony, rather than any experience with tones, changes how tones impact categorization is to directly compare the Synchronous and Asynchronous Tone conditions. Numerically, infants in the Synchronous Tone condition showed a greater familiarity preference than those in the Asynchronous Tone condition. The effect was moderate in size and in the predicted direction, and the Bayes Factors provided anecdotal evidence that groups differed (BFalt = 1.148, BFnull =.871). However, a directional *t* test comparing the conditions did not reach significance; *t* (39) = 1.377, *p* = .088, *d* = .43, observed power = .385. Thus, although we found evidence that a history of synchrony with object motion changed the effects of tones on categorization, making them act like words, the evidence that this was an effect of synchrony *per se* is weaker. Also note that neither the test performance of the Synchronous or Asynchronous condition differed from chance (see Table 1). Furthermore 12 of 20 infants (60%) in the Synchronous group and 8 of 21 in the Asynchronous group (38%) showed a familiarity preference, and binomial tests revealed that neither rate differed from chance (*p* = .5 and *p* = .383 respectively. Thus, overall the effects are of a similar, moderate strength to those in Experiment 1.

In Experiment 1 we found potential evidence that infants in the Word condition were more attentive during the categorization training than infants in the Tone condition, and the former also showed a greater familiarity preference at test. If that difference in test performance is primarily due to differences in attentiveness, rather than to effects of the signal on categorization, then we might likewise expect infants in the Synchronous Tone condition to have been more attentive than infants in the Tone condition. Infants in the Synchronous Tone condition spent on average 54% of the Categorization Training attending to the display as compared to 48% in the Tone condition, but there was no evidence that this was a meaningful difference: A directional *t* test yielded no significant difference (*t* (34) = 1.032, *p* = .155 *d* = .346, observed power = .289), and the Bayesian *t* test provided anecdotal support for the hypothesis that they did not differ (BFalt = .797 and BFnull = 1.22). Thus, there is no evidence that the differences in the Tone and Synchronous Tone groups’ categorization performance were due to differences in attention during the Categorization Training phase.

Likewise, the Synchronous Tone group was not more attentive than the Asynchronous Tone group, who attended for 56% of the time; directional *t* (39) = .355, *p* = .638, *d* = .111, observed power = .098; BFalt = .242 and BFnull = 4.125. There were no correlations between the attentiveness during the categorization Training and Test performance in the Synchronous Tone condition (*r* = -.210, *p* = .375) in the Asynchronous Tone condition (*r* = -.166, *p* = .616) or in the two groups combined (*r* = -.144, *p* = .368). Altogether, these results suggest that it is unlikely that group differences in test performance can be attributed to differences in attention during the Training phase (i.e., the difference found between infants in the Synchronous Tone and Tone conditions at Test was not accompanied by a difference in attention during training, and there was no other evidence that attention during Training was linked to test performance).

There were also no differences in the proportion of time that infants in the Synchronous Tone and Asynchronous Tone conditions attended during the Pretraining Phase (*M* = .849, *SE* = .021 and *M* = .836, *SE* = .026 respectively; independent samples *t* (39) = -.358, *p* = .722, d = .112, observed power = .064; BFalt = .322 and BFnull = 3.105. In fact, the value of the BFnull suggests moderate evidence that they didn’t differ. Moreover, there were no correlations between the amount of time spent attending during Pretraining and test preferences for either the Synchronous or Asynchronous Tone groups (*r* = .004, *p* = .987 and *r* = -.011, *p* = .961 respectively) or for the two groups combined (*r* = -.018, *p* = .913).

**Discussion**

In Experiment 1 we replicated previous findings that words and tones have different effects on categorization. In Experiment 2 we were able to change how tones affect categorization by giving infants a Pretraining phase in which tones occurred in synchrony with object motion. After this pretraining, infants subsequently performed differently from those who had no prior experience with tones, and equivalently to those hearing words. Infants given pretraining with tones that occurred out of synchrony with object motion did not differ from the Tone group, however, they also did not perform differently from the Synchronous Tone group on the categorization task. Thus, the evidence that synchrony is responsible for changes in how a signal influences categorization is mixed. We attempted to provide stronger evidence on the role of synchrony in object categorization in Experiment 3.

**Experiment 3**

Our central hypothesis is that cross-modal synchrony, like that present when caregivers name objects as they show them to infants, contributes to the special effects that words have on object categorization. In Experiment 2 we found evidence that experience with synchrony between tonesand object motion can cause tones to have effects that are similar to those of words. These results suggest that synchrony can lead a signal to have the effects that words do, and thus that word-object synchrony may contribute to the distinctive effects that words have on categorization. However, stronger evidence for this hypothesis would come from a direct manipulation of word-object synchrony.

By 3 months of age, infants have likely accumulated substantial experience with word-object synchrony, and thus we cannot realistically eliminate or even systematically control their experience with it prior to the experiment. And, because words already support categorization by 3 months, we can’t determine whether synchrony leads words to have these effects *de novo*. However, we can manipulate the extent to which words and objects have an immediate, salient history of synchrony within the experiment by exposing infants to a Pretraining phase in which words and objects occur in or out of synchrony, as we did with tone-object synchrony in Experiment 2, and then test whether words have an even stronger effect on object categorization than that typically observed.

To do so, we built on evidence that there is a developmental shift in the effects of words on categorization performance across 3- to 6-months of age. At 3-months infants show a familiarity preference at test when words are paired with exemplars during the familiarization phase (in our Experiment 1 and in Ferry et al. 2010), but 4-9-month-olds show a novelty preference under identical conditions (Ferry et al., 2010; Fulkerson and Waxman, 2007). This shift from a familiarity preference to a novelty preference is consistent with evidence that, all other things being equal, older infants are more likely to show a novelty preference in VPC tasks (Hunter & Ames, 1988). A familiarity preference is also generally a sign of weaker representations and poorer recognition of familiarized objects or events. For example, in a series of experiments using a VPC task, Bahrick and colleagues (Bahrick & Pickens, 1995; Bahrick et al., 1997) found that 3-month-olds familiarized to a video of a moving object showed a novelty preference when tested with that object paired with a novel one after a one-minute delay, but a familiarity preference when tested after a month-long delay, consistent with weaker recognition of the familiarized object. Critically, the familiarity preference at the month-long delay could be shifted to a novelty preference by reactivating the representation of the object prior to test (i.e., by briefly refamiliarizing them to the object on the day prior to testing). It seems that by reactivating their memory for the familiarized item, infants were better able to access it at test, and thus showed more robust recognition (i.e., a novelty preference).

In the studies by Ferry et al. (2010) and Fulkerson and Waxman (2007), the shift in preference across 3 to 4 months suggests that the older infants more quickly and strongly grouped the exemplars into a category during the familiarization phase, and/or were more likely to perceive the novel exemplar from the familiarized category as similar at test, relative to younger infants. Indeed, infants’ representations become more flexible with age, such that infants who were conditioned to kick to activate a mobile could subsequently have their memory for the mobile reactivated by a similar but not identical mobile by 9 months, but only the identical mobile consistently serves as a memory cue at earlier ages (Hartshorn et al., 1997).

The fact that infants still fail to form a category at 4 months when the exemplars are paired with tones at 4 months (Ferry et al., 2010) suggests that this change in performance is not due to general improvements in the ability to form visual categories. Rather, words appear to have a more powerful effect on categorization in older infants by supporting better exemplar encoding, stronger similarity-based grouping of those exemplars, better recognition of the novel exemplar from the familiarized category, or (likely) a combination of all of these. If accruing experience with word-object synchrony contributes to this shift, then concentrated experience with word-object synchrony in the lab could boost 3-month-olds’ performance to the level of older infants (i.e., shifting them from a familiarity to a novelty preference). This possibility is consistent with evidence that while encoding amodal information can came at the cost of encoding unimodal information early in development due to limited attentional resources, the presence of synchrony can eventually lead to enhanced encoding of unimodal information: As amodal information is more readily encoded, more resources become available for unimodal processing (Bahrick & Lickliter, 2014). In the case of word-object synchrony, this might mean that infants have more resources available for encoding the visual properties of objects that are relevant for categorization, leading to a novelty preference at test.

**Methods**

**Participants.** Participants were 40 monolingual, English-learning infants, 19 of whom were female, and who were on average 3 months and 15 days old (the range was 3;1 to 3; 29). We used the same sample sizes as in Experiments 1 and 2 to promote comparability. There were 35 European American participants, 2 Black, and 3 mixed race, according to parent report. As in Experiments 1and 2, infants were born full-term, had normal hearing and no history of chronic ear infections or serious health or cognitive issues. An additional 3 infants were tested, but their data were excluded because of equipment failure (2) and experimenter error (1). No infants needed to be excluded for fussiness or inattention.

**Materials.** The materials were largely identical to those from Experiments 1 and 2, except that in the Pretraining Phase the sine-wave tones were replaced with a word (“vimo”, which was not used during the Categorization Task). The materials for the Categorization Task consisted of the Training and Test phases from the Word condition of Experiment 1.

**Procedure.** All infants participated in the Pretraining phase, viewing videos in which object motion and words were either synchronous or asynchronous (the Synchronous Word and Asynchronous Word conditions, respectively). Infants were then given the Categorization Training and Test, as in the Word condition from Experiment 1.

**Results**

Reliability, calculated as in the other two experiments, was 97%, and the same exclusion criteria were applied. Figure 2 shows performance in the two conditions in this experiment, alongside the data from Experiments 1 and 2 for comparison. As predicted, infants in the Synchronous Word condition showed a larger novelty preference than those in the Word condition from Experiment 1 (directional *t* (41) = 1.80, *p* = .040, *d* = .549, observed power = .524), with a medium effect size. A Bayesian *t* test indicated that the evidence in favor of accepting the alternative hypothesis was anecdotal (BFalt = 2.036, and BFnull = .491). This difference did not appear to result from a general effect of prior experience with words and objects, as infants in the Asynchronous Word condition did not show a greater novelty preference than those in the Word condition (directional *t* (39) = .454, *p* = .655, *d* = .126, observed power = .105), with moderate evidence that these groups did not differ (BFalt = .236 and BFnull = 4.236). Likewise, and potentially even more critically, infants in the Synchronous Word condition showed a larger novelty preference than those in the Asynchronous Word condition (directional *t* (38) = 1.996, *p* = .027, *d* = .632, observed power = .625), with a medium effect size and anecdotal evidence that the groups differed (BFalt = 2.773 and BFnull = .361).

Altogether these data suggest that training with word-object synchrony alone shifted 3-month-olds towards the more mature novelty preference typically not shown until 4 months of age. And, this effect was seen only in infants given pretraining with words occurring in synchrony with object motion. Note that, similar to Experiments 1 and 2, the performance of infants in neither the Synchronous or Asynchronous Word condition differed from chance (*t*s < 1.42, *p* > .170; see also Table 1). We found that 14 of 21 infants (67%) in the Synchronous Word condition showed a novelty preference, while only 6 of 19 (32%) in the Asynchronous word condition did, but these rates did not differ from chance in a binomial test (*p* = .189 and *p* = .167 respectively).

If this boost in categorization is driven by differences in attentiveness during the categorization Training, rather than their history of synchrony with object motion, then we should find evidence that infants in the Synchronous Word condition were more attentive than those in the Asynchronous Word condition during Training. However, there were no differences in how much infants in the Synchronous and Asynchronous Word conditions attended (*M* = .580, *SE* = .028 and *M* = .515, *SE* = .048, respectively; *t* (38) = 1.214, *p* = .232; BFalt = .553 and BFnull = 1.809). There were no correlations between attentiveness and test preferences for either condition (*r* = .008, p = .971 in the Synchronous Word condition, and *r* = -.075, *p* = .760 in the Asynchronous Word condition) or when the groups were combined (*r* = .021, *p* = .896).

We also tested whether infants in the Synchronous Word condition could have showed a more mature signature of categorization because of increased attentiveness during the Pretraining phase. Interestingly, although both groups were highly attentive, infants in the Asynchronous Word condition were more attentive to the Pre-Training videos than those in the Synchronous Word condition (*M* = .920, *SE* = .015, and *M* = .853, *SE* = .018, respectively; *t* (38) = 2.749, *p* = .008, d = .871, observed power = .766), BFalt = 5.344 and BFnull = .187. This result was unexpected given that synchrony typically attracts infants’ attention, at least in events with shorter durations, and also because there were no differences in how much infants payed attention to the training events when *tones* were synchronous vs. asynchronous with object motion. They may have payed equal attention to the pretraining videos featuring tones because they had no learned expectation about how tones and objects co-occur. In contrast, infants are likely to have experienced word-object synchrony, and thus the events in the Asynchronous pretraining videos could have been unexpected and therefore attention-getting. Importantly, whatever the reason for this difference, it is unlikely that attention differences during Pretraining can explain the differences in the two groups’ categorization performance: If more exposure to some generic aspect of the Pretraining videos, such as to words and/or objects, rather than the synchrony between them, was central to the changes in categorization performance relative to the Word condition from Experiment 1, then the Asynchronous condition should have flipped to novelty preference. Instead, only the Synchronous Word group changed their preference. Furthermore, there was no correlation between the amount of time spent attending during Pretraining and test performance (*r* = .184, *p* = .424 in the Synchronous Word condition, and *r* = .036, *p* = .882 in the Asynchronous Word condition) or when the groups were combined (*r* = -.02, *p* = .903). Altogether then, there is no evidence to suggest that differences in attention during Pretraining or the categorization Training, rather than the manipulation of synchrony, could have led to the observed differences in test performance.

**Discussion**

Giving 3-month-olds concentrated experience with word-object synchrony changed how words subsequently impacted their performance on a categorization task. While 3-month-olds typically show a familiarity preference in categorization tasks when exemplars are accompanied by words (Experiment 1; Ferry et al., 2010), we were able to lead them to show a novelty preference (characteristic of older infants) by giving them a concentrated dose of word-object synchrony prior to the categorization task. This effect was specific to prior experience with word-object synchrony, as categorization performance in the Synchronous Word condition differed from those in the Asynchronous condition. We discuss potential mechanisms by which experience with word-object synchrony impacts categorization in the General Discussion.

**General Discussion**

In Experiment 1 we replicated prior findings that words and tones have different effects on infants’ ability to form visually-based categories by 3 months of age. In Experiment 2 we found that tones can come to act like words when they have a recent history of occurring in synchrony with object motion. In Experiment 3 we found that giving infants concentrated experience with word-object synchrony can also impact the effects of words on categorization, leading 3-month-olds to perform like 4-month-olds (Ferry et al., 2010). Moreover, only experience with synchrony between words and objects and between tones and objects had these effects; experience with tones or words that occurred asynchronously with object motion did not affect subsequent categorization. Altogether, these results support the hypothesis that synchrony between words and object-demonstration gestures, common in the act of naming, contribute to the special effects that words have on how infants learn about and categorize objects.

However, there are some details of our results that bear further scrutiny. First, although we replicated previous findings that words and tones lead to different effects on categorization, the effect size was smaller than in previous studies, and dissipated within 5 seconds. This suggests that the strength of learning in our categorization task was weaker and shorter-lived than that observed in prior work, possibly due to differences in our materials or experimental set-up. Likewise, we did not find strong evidence of successful categorization for any of the individual groups across the 3 experiments. Nonetheless, our data clearly show both that words and tones have different effects on categorization by 3 months, and that these effects can be both elicited by and enhanced by synchrony. By showing that the effects of words on categorization can be moderate in size, and can be short lived, our data provide an important contribution to the literature. They also suggest that in future work with this paradigm, more modest effect sizes should be used to calculate necessary sample sizes.

Another issue concerns the performance of infants in the Tone condition in Experiment 1, whose preference for the novel exemplar did not differ from chance, but the effect size of that preference was larger than that of the Word group’s familiarity preference. It is possible that infants in the Tone condition learned something about the training exemplars that allowed them to discriminate between the exemplars from the novel and familiar categories at test, and led them to show a novelty preference. However, there are several previous demonstrations that tones do not facilitate infants’ categorization (i.e., Balaban & Waxman, 1997; Ferry et al., 2010; Fulkerson & Waxman, 2007). Given that words and tones have been shown to influence categorization performance differently in several other studies, our findings are likely to reflect a real underlying difference between the conditions, whether infants in the Tone condition were simply failing to categorize, or whether they were doing something systematically different. Regardless, tone-object synchrony had an effect on how tones subsequently impacted categorization. Furthermore, in Experiment 3 we found that synchrony can amplify the effects of words on categorization, both relative to no prior experience and to prior experience with words and objects occurring asynchronously. Thus, our results clearly suggest that synchrony influences the extent to which both tones and words affected infants’ categorization.

Note also that our results are consistent with evidence that tones can mimic the facilitatory effects of words in a categorization task, given the right conditions. For example, Ferguson and Waxman (2016) found that tones facilitated categorization after being used in a communicative exchange in which one person spoke using the infants’ native language and another beeped back using tones. Our results are consistent, in that synchrony is prevalent in communicative interactions. However, the experience that infants were given with tones in our experiment did not contain people, nor were the tones used in alternation with speech, which may have served to explicitly equate them. Instead, infants were exposed to timing regularities that are spontaneously generated in human communication, but embedded in a situation in which other features of communicative exchanges (such as the presence of people, eye contact, and speech) were absent. These findings suggest that the amodal temporal structure of communicative behaviors plays a critical role in the special effects of words in early object cognition. Moreover, we show that this redundancy can have a more fundamental effect on how infants learn: Systematic experience with sound-object synchrony affected how infants later processed those sounds and objects, even when they no longer exhibited synchrony, suggesting that synchrony can have a lasting impact on how infants encode and learn about the world around them.

But how exactly did synchrony impact subsequent categorization? One possibility is that a history of sound-object synchrony leads infants to pay more attention to objects in the context of those sounds, and thus to encode them better, even when synchrony is absent. Across the experiments there was no evidence that synchrony led to gross increases in attention, either to events involving audio-visual synchrony, or to signals that had been involved in such synchrony. However, differences in attention are not always revealed by differences in looking time, and there is evidence from ERP studies that synchrony can leads to greater attention and encoding (Reynolds et al., 2014).

Another possibility, which is not inconsistent with the possibility that increases in attention are relevant to the effects of synchrony on categorization, is that synchrony influenced *what* infants encoded. Indeed, previous work suggests that the presence of amodal structure, while leading to enhanced encoding of redundant information, can lead to impoverished encoding of information that is specified in just one modality (Bahrick, Flom, & Lickliter, 2002). (Bahrick, Licklier, & Flom, 2006). Thus, it is possible that a history of sound-object synchrony led infants to subsequently attend to *sound-object co-occurrence* in the categorization task, and thus to encode the objects more poorly. It seems paradoxical that poor object encoding would improve categorization. However, rather than leading to uniformly impoverished representations, the presence of amodal structure may lead to representations in which some information is relatively spared. For example, shape is often privileged in infants’ object representations when encoding is difficult (Wilcox, 1999), and thus synchrony may lead to relatively robust encoding of object shape in comparison to more fine-grained features, in turn leading to improved categorization.

Another possibility is that synchrony leads to stronger, or more precise and detailed, object encoding. This possibility is consistent with evidence that there is a salience hierarchy in the encoding of multimodal events involving synchrony: Audio-visual synchrony, while initially serving to unify information at the expense of unimodal encoding, eventually scaffolds learning about unimodal features of events, as infants gain experience with processing amodal relations (Bahrick, 2001; Bahrick et al, 2010). In the current experiment, the more experience infants have with sound-object synchrony, the more readily they may encode object properties in the presence of those sounds, and the more strongly those effects would influence encoding even when synchrony is absent. Infants are likely to have very little experience with tone-object synchrony prior to the experiment, so effects of their experience with it during Pretraining phase on subsequent object encoding during the categorization Training and Test in Experiment 2 would be weak. Likewise, at 3 months infants would have some experience with word-object synchrony, and these effects of words may only weakly carry over to nonsynchronous experience with sounds and objects during the categorization training in Experiment 1. In both of these cases, object encoding in the categorization Training would get a boost, but it would likely be relative weak, leading to a familiarity preference at test. Giving 3 month-olds concentrated experience with word-object synchrony in Experiment 3 may have led words to have a more powerful effect on exemplar encoding during the categorization training, and to a novelty preference. These possibilities are speculative, and more work should be done to directly investigate the effects of word-object synchrony on encoding words, and how such effects carry over to encoding words and objects when synchrony is absent.

Our findings also raise the question of whether synchrony changes the nature of the connection between words and objects in infants’ minds. A fundamental question about language learning is how words come to refer, or to “mean” something. Co-occurrence between word and objects alone cannot specify the nature of the underlying connection between them (i.e., it could be symbolic, causal, etc.). An intriguing possibility, and one we are testing in follow-up experiments, is that synchrony contributes to these effects. Specifically, we hypothesize that by promoting the unification of word and objects, synchrony may contribute to the establishment of reference, over and above association.

In sum, we found that sound-object synchrony, which is characteristic of caregiver communication about objects, can result in the special effects that words have on infants’ categorization. Words may have these effects on categorization in the absence of any experience, but our results show that it is possible for a signal to take on these effects solely by virtue of this kind of synchrony. Thus, it is possible that synchrony is a factor (though surely not the only factor) in the effects that words have on infant cognition, and in the origins of symbolic reference.

**References**

Althaus, N., & Mareschal, D. (2014). Labels direct infants’ attention to commonalities during novel category learning. *PloS One*, *9*, e99670. DOI: 10.1371/journal.pone.0099670

Bahrick, L. E., Hernandez-Reif, M., & Pickens, J. N. (1997). The effect of retrieval cues on visual preferences and memory in infancy: Evidence for a four-phase attention function. *Journal of Experimental Child Psychology*, *67*(1), 1-20. https://doi.org/10.1006/jecp.1997.2399

Bahrick, L. E., & Pickens, J. N. (1995). Infant memory for object motion across a period of three months: Implications for a four-phase attention function. *Journal of Experimental Child Psychology*, *59*(3), 343-371. https://doi.org/10.1006/jecp.1995.1017

Bahrick, L. E., & Lickliter, R. (2014). Learning to attend selectively: The dual role of intersensory redundancy. *Current Directions in Psychological Science*, *23*, 414-420. [doi.org/10.1177/0963721414549187](https://doi.org/10.1177%2F0963721414549187)

Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3‐month‐old infants. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *41*, 352-363. <https://doi.org/10.1002/dev.10049>

Bahrick, L. E., Lickliter, R., Castellanos, I., & Vaillant‐Molina, M. (2010). Increasing task difficulty enhances effects of intersensory redundancy: Testing a new prediction of the intersensory redundancy hypothesis. *Developmental Science*, *13*(5), 731-737. https://doi.org/10.1111/j.1467-7687.2009.00928.x

Bahrick, L. E., Lickliter, R., & Flom, R. (2006). Up versus down: the role of intersensory redundancy in the development of infants' sensitivity to the orientation of moving objects. *Infancy*, *9*, 73-96. <https://doi.org/10.1207/s15327078in0901_4>

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, *64*, 3-26. <https://doi.org/10.1006/jecp.1996.2332>

Baldwin, D. A., & Markman, E. M. (1989). Establishing word-object relations: A first step. *Child Development*, 381-398. DOI: 10.2307/1130984

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*, 115. [https://doi.org/10.1037/0033-295X.94.2.115](https://psycnet.apa.org/doi/10.1037/0033-295X.94.2.115)

Booth, A. E., & Waxman, S. R. (2003). Mapping words to the world in infancy: Infants' expectations for count nouns and adjectives*. Journal of Cognition and Development, 4*, 357-381. <https://doi.org/10.1207/S15327647JCD0403_06>

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Ferguson, B., & Waxman, S. R. (2016). What the [beep]? Six-month-olds link novel communicative signals to meaning. *Cognition*, *146*, 185-189. <https://doi.org/10.1016/j.cognition.2015.09.020>

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental Psycholinguistics: On-Line Methods in Children’s Language Processing*, *44*, 97.

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*, 472-479. <https://doi.org/10.1111/j.1467-8624.2009.01408.x>

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*, 15231-15235. https://doi.org/10.1073/pnas.1221166110

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, *43*, 238. <https://doi.org/10.1037/0012-1649.43.1.238>

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, *105*, 218-228. <https://doi.org/10.1016/j.cognition.2006.09.005>

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*, 878-894. <https://doi.org/10.1111/1467-8624.00197>

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6-to 8-month-old infants and learning of word–object relations. *Infancy*, *9*(3), 259-288. DOI: [10.1207/s15327078in0903\_1](https://doi.org/10.1207/s15327078in0903_1)

Gogate, L. J., Prince, C. G., & Matatyaho, D. J. (2009). Two-month-old infants' sensitivity to changes in arbitrary syllable–object pairings: The role of temporal synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 508. [https://doi.org/10.1037/a0013623](https://psycnet.apa.org/doi/10.1037/a0013623)

Hartshorn, K., Rovee‐Collier, C., Gerhardstein, P., Bhatt, R. S., Klein, P. J., Aaron, F., ... & Wurtzel, N. (1998). Developmental changes in the specificity of memory over the first year of life. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *33*(1), 61-78. https://doi.org/10.1002/(sici)1098-2302(199807)33:1<61::aid-dev6>3.0.co;2-q

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, 5, *69-95.*

Iverson, J. M., Longobardi, E., Spampinato, K., & Cristina Caselli, M. (2006). Gesture and speech in maternal input to children with Down's syndrome. *International Journal of Language & Communication Disorders*, *41*(3), 235-251. DOI: [10.1080/13682820500312151](https://doi.org/10.1080/13682820500312151)

JASP Team (2019). JASP (Version 0.11.1)[Computer software].

Kelso, J. S., Tuller, B., & Harris, K. S. (1983). A “dynamic pattern” perspective on the control and coordination of movement. In *The Production of Speech* (pp. 137-173). Springer, New York, NY.

LaTourrette, A., & Waxman, S. R. (2019). A little labeling goes a long way: Semi‐supervised learning in infancy. *Developmental Science*, *22*, e12736. <https://doi.org/10.1111/desc.12736>

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lickliter, R., Bahrick, L. E., & Markham, R. G. (2006). Intersensory redundancy educates selective attention in bobwhite quail embryos. *Developmental Science*, *9*, 604-615. <https://doi.org/10.1111/j.1467-7687.2006.00539.x>

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago press.

Reynolds, G. D., Bahrick, L. E., Lickliter, R., & Guy, M. W. (2014). Neural correlates of intersensory processing in 5‐month‐old infants. *Developmental Psychobiology*, *56*(3), 355-372. DOI 10.1002/dev.21104

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: differentiation of multimodal information. *Psychological Bulletin*, *121*, 437. [https://doi.org/10.1037/0033-2909.121.3.437](https://psycnet.apa.org/doi/10.1037/0033-2909.121.3.437)

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*, 217-242. <https://doi.org/10.1006/cogp.2001.0764>

Waxman, S., & Booth, A. (2003). The origins and evolution of links between word learning and conceptual organization: New evidence from 11‐month‐olds. *Developmental Science*, *6*, 128-135. <https://doi.org/10.1111/1467-7687.00262>

Wagenmakers E-J, Marsman M, Jamil T, Ly A, Verhagen J, Love J, et al. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications*. Psychonomic Bulletin & Review, 25(1)*, 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wilcox, T. (1999). Object individuation: Infants’ use of shape, size, pattern, and color. *Cognition*, *72*, 125-166. <https://doi.org/10.1016/S0010-0277(99)00035-9>

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1-34.

1. Measures of performance in the second 5 seconds are also reported in Table 1; note that the patterns are either consistent or diminish across windows, as in Experiment 1. [↑](#footnote-ref-1)