The effect of emotional prosody and referent characteristics on novel noun learning

Melissa K. Jungers[1], Julie M. Hupp[1], Jarrett A. Rardon[2], Samantha A. McDonald[1] and Yujin Song[3]

[1]Department of Psychology, The Ohio State University at Newark, Newark, OH, USA; [2]Nationwide Children's Hospital, Columbus, OH, USA and [3]Department of Linguistics, The Ohio State University at Columbus, Columbus, OH, USA
Corresponding author: Melissa K. Jungers; E-mail: Jungers.2@osu.edu

Abstract

Prosody includes the pitch, timing and loudness in speech, which can convey meaning and emotion. This study examines whether prosodic categories affect novel noun learning and whether the referent characteristic influence learning. Previous research showed that emotional prosody interfered with adults' noun learning (West et al., 2017), but it had no effect on children (West et al., 2022). However, these researchers varied their method across ages, including animacy and complexity of the referent, and it is unclear if the results extend beyond the three emotional prosodies tested. Participants in the current set of studies heard novel words presented in five prosodic categories (within-subject) in order to learn the label for either animate or inanimate objects (between-subject). Study 1 compared inanimate objects and aliens, with better noun learning performance for inanimate objects. Study 2 compared inanimate objects with the same objects with faces added, but there was no difference in noun learning by object type. Both studies showed differences in noun learning by the prosodic category, with warning less accurate than naming. These results demonstrate how extralinguistic factors like prosody, attention and referent complexity influence noun learning.

1. Introduction

A speaker's message is conveyed not only by *what* is said, but also by *how* it is said. Acoustic variations in speech, such as timing, stress and pitch, are known as prosody (Cutler et al., 1997; Warren, 1997). Prosody helps listeners interpret syntactically

---

ambiguous sentences (Lehiste, 1973; Lehiste et al., 1976) and guides visual search for referents based on emphasis or pitch accent (Ito & Speer, 2008).

Prosody also works analogically, with changes in voice corresponding to changes in a visual scene. For example, speakers use pitch to indicate brightness and size (Marks, 1987) and speech rate to indicate the speed of an object, with faster speech corresponding to faster moving objects (Shintel et al., 2006). The size of objects is indicated by pitch and speed, even when the words are nonsense, with slower rates and lower pitch for larger objects, but faster rates and higher pitch for smaller objects (Nygaard et al., 2009). Adults and children both produce and are sensitive to these acoustic analogs, connecting faster moving objects with faster speech (Hupp & Jungers, 2013). Similarly, adults select still images that imply motion, such as a running horse, when listening to faster speech (Shintel & Nusbaum, 2007).

In addition to providing visuospatial information, prosody also indicates a speaker's intention. For example, listeners produced and recognized speech acts such as criticism, doubt, naming, suggestion, warning and wish from both a single word and even from a nonword regardless of valence or arousal (Hellbernd & Sammler, 2016). Likewise, adults and preschool children identified an intended referent based on prosodic intentions (warn, doubt, name) for both words and nonwords (Hupp et al., 2021). Children are sensitive to prosodic intention, selecting a broken toy over an intact toy in response to a negative-sounding voice (Berman et al., 2010). Children's eye gaze also reflects understanding prosodic intention as they look toward a broken or special toy based on the negative or positive voice (Berman et al., 2013). Prosodic cues help listeners identify the referent in a visual scene and understand a speaker's intention. However, how do these cues affect word learning in adults?

Word learning can be enhanced by the speaker's tone of voice when this prosodic cue adds relevant information. When adult listeners heard novel adjectives spoken with a prosody that indicated one part of a contrasting pair of antonyms (big-small) while looking at two images of the same contrast (elephant-ant), they demonstrated adjective learning consistent with the prosody and could generalize this learning to a new referent even when a neutral tone of voice was used in the test (Reinisch et al., 2013). Similarly, prosodically conveyed semantic information influences how listeners learn a novel adjective for a referent in an antonym pair (Shintel et al., 2014). Adults viewed picture pairs (big dog-small dog) and learned novel adjectives spoken with either congruent or incongruent prosody. Their memory for the novel adjectives was not affected by congruency when tested immediately, but 24 hours later, the congruent words were better remembered (Shintel et al., 2014).

Word processing can also be affected by prosodic cues related to emotion. Emotion can be conveyed prosodically through changes in fundamental frequency or speech rate (Lieberman & Michaels, 1962; Mauchand & Pell, 2021; Murray & Arnott, 1993), and emotional prosody is processed in different brain regions than neutral prosody (Lei et al., 2021). For example, doubt is recognized with a long stimulus and rising pitch contour, warning is recognized by high mean pitch and intensity, and naming is marked by a short stimulus, low mean pitch and falling pitch contour (Hellbernd & Sammler, 2016). Prosody and semantics (word meaning) can work together to help listeners identify an emotional intent, but prosody dominates in a task where they conflict (Ben-David et al., 2016). There is an attentional negativity bias to identify emotionally neutral words spoken with emotional prosody

such that identification is slower for words spoken with happy or sad prosody relative to neutral prosody (Krestar & McLennan, 2019).

Emotion also affects word processing in print. Participants respond more quickly to printed positive and negative words than neutral words in a word/nonword identification task, even when arousal is held constant (Kousta et al., 2009). This faster processing of emotionally significant stimuli could lead to more accurate word learning. In a word-learning task, negative words were better remembered than neutral words (Kensinger & Corkin, 2003), and others have shown that novel words learned with a positive connotation have a retrieval advantage (Snefjella et al., 2020). There may be a memory advantage for emotion words over neutral words, but the words in the previous studies were presented as written text. How does hearing words with emotional prosody affect word learning?

West et al. (2017) tested adults for their learning of three-syllable nonsense words presented with a happy, neutral or fearful prosody. The stimuli were created by a trained actress, and the stimuli were judged for their fit to the prosodic categories on a Likert-type scale by independent listeners. The words were paired with a visual image of an alien, and the participants engaged in five repetitions of learning and recall. All participants better recalled neutral over fearful labels. Those with lower autism-like traits also learned neutral noun labels better than happy noun labels. This study suggests that emotion interferes with effective word learning in typical adults. Emotion is not relevant to the process of word learning and served as a distraction, which contrasts with previous work in which there was a memory advantage for printed emotion words (Kensinger & Corkin, 2003; Snefjella et al., 2020).

In a similar study with 7- to 9-year-old children, nonsense words were paired with novel objects (West et al., 2022). Children watched a video of a female speaker who pronounced the words with happy, fearful or neutral prosody within a sentence context while holding a picture of the object. Unlike adults, children showed no difference in memory for noun labels by emotional prosody. West et al. (2022) claim that the differences in prosody's influence on word learning with adults and children across these two studies reflect the development of the processing of extraneous emotional information that may arise after the age of 9.

The goal of the current study is to examine adults' word learning with emotional prosody by replicating and expanding the work of West et al. (2017, 2022). In their adult study, the emotion of the prosody (happy, fearful neutral) affected word learning, with better recall for neutral than fearful items across all participants and better recall for neutral than happy items for those with lower autism-like traits. The fearful condition was also the slowest and least accurate prosody in the adult recognition task. The child study showed no differences in word recall or recognition by emotion. However, there were methodological differences between the two studies that make it difficult to directly compare their findings. The adults learned 30 individual words and 30 static pictures of aliens and were tested for recall after each of the five learning blocks. They also performed a recognition task that included printed nonsense words. The children learned nine novel labels while watching videos with a speaker producing full sentences to describe the pictured novel objects. After each of the two learning blocks, the children performed a picture recognition task with written and verbal instruction to find the target item, and then they performed one recall test. Unlike the adults, the children were then tested for their ability to generalize to similar objects.

In addition to these procedural differences across ages, the two West et al. (2017, 2022) studies also differed in their labeled referents (Aliens, Objects). For adults, the referents were complex, animate aliens, with unique coloring, number of limbs and clothing, but for children, the referents were simple, inanimate, toy-like objects. A developmental change could account for the different findings across studies, as suggested by West et al. (2022), but in addition to the aforementioned procedural differences, the stimuli themselves differed across age groups in animacy and complexity, which may also affect noun learning. There is evidence that animacy can aid learning of single words, words in lists and pictures (Aka et al., 2020; Bonin et al., 2014; Meinhardt et al., 2019; Nairne et al., 2013; Popp & Serra, 2016) and simple objects are easier to remember than complex objects (Eng et al., 2005; Mishra, 1984).

Furthermore, the West et al. (2017, 2022) studies included only neutral, fearful and happy prosodies, but there is evidence that adults and children identify speech acts such as doubt and warn as well (Hellbernd & Sammler, 2016; Hupp et al., 2021). The current study expands the number of prosodies to include Doubt, Fear, Happy, Name (Neutral) and Warn. The additional speech acts in the current study will further elucidate the role of emotion in word learning. To better understand the effects of West et al.'s research, the current study tests adults' novel noun learning across a variety of prosodies and across both animate and inanimate referents, using referents similar to those of West et al. (2017, 2022).

2. Study 1: Aliens versus objects

The goal of Study 1 was to measure adults' novel noun learning for animate aliens and inanimate objects across five prosodic conditions. Specifically, are there differences across emotional prosody conditions in word learning? Also, does the type of referent matter? How does learning progress over time? Based on prior research (West et al., 2017), it is anticipated that there will be differences in word learning by prosody. In addition, there will be improved word learning across time.

2.1. Method

2.1.1. Participants

The participants included 238 undergraduate students from a regional campus of a large university who participated for research credit in their introductory psychology course. Participants were excluded from the analyses for a variety of reasons: failure to complete the entire session ($n = 36$), non-native English speakers ($n = 15$), uncorrected hearing or vision problems ($n = 4$) or below 80% on catch trials ($n = 23$), for a final sample of 160 (M age = 19.08 years, SD = 3.03). There were 71 males (44%), 83 females (51%), 3 nonbinary, gender-fluid or agender (2%) and 3 undisclosed (2%). The Institutional Review Board at the authors' home institution granted ethics approval for this project, and the participants provided informed consent prior to participation.

2.1.2. Materials

2.1.2.1. Pilot study.
An initial pilot study was conducted online using Qualtrics with undergraduate students drawn from the same participant pool ($n = 49$), who were not in the main study, to verify prosodic categories (Doubt, Fear, Happy, Name, Warn) of

15 novel labels (e.g., *tebos*) recorded by a female speaker. Participants heard 125 items, which were a random subset of 41 possible novel nouns (from Gupta et al., 2004) spoken in isolation with one of the five prosodic categories, and a set of catch trials indicating the correct answer to ensure they were paying attention. The participants were instructed to select which of the five prosodies was being used for each item. In the catch trials, there was no word; instead, there was a spoken sentence with instructions (e.g., 'For this one, please select Fear'). Participants who missed more than 20% of the catch trials were not included in the analyses. Accuracy for the intended prosody was calculated for each individual item. Any words with accuracy below 80% for any individual prosody were excluded entirely. Then, the 15 words with the highest accuracy across all 5 prosodies were chosen for use in this study, with an average accuracy score of 86.88%. Thus, the final set had 75 items; 15 words were produced with 5 prosodies. All data, materials and analyses used in this research are available as open access at the following link: https://osf.io/jgbm6/?view_only= a0c3331984e4482da6a9a6bf64d0de9e.

2.1.2.2. Acoustic details. In addition to the behavioral pilot study, the final 75 items were also analyzed acoustically. First, the stimuli in each emotion category were analyzed for pitch contour. Happy, Fear and Warn emotions have a similar pitch pattern (contour); however, the f0 ranges of the rise and fall are greater and their slopes are steeper for the Happy stimuli than for the other two (especially compared to Fear). The Name category has the smallest f0 range. Doubt is unlike the other emotions because it ends with a rise (see Figure 1 for a sample word).

Next, acoustic measures of prosodic features, including duration, f0 measures and intensity, were compared using separate linear mixed-effects models. For word duration, there was a main effect of emotional prosody category, $F (4, 60) = 67.37$, $p < .001$. The post hoc pairwise comparisons revealed that the durations of all pairs of emotional prosody categories were significantly different from each other ($p < .05$), except between Name and Fear and Doubt and Fear. Overall, Warn had the shortest duration relative to the other emotional prosody categories. For mean f0, there was a
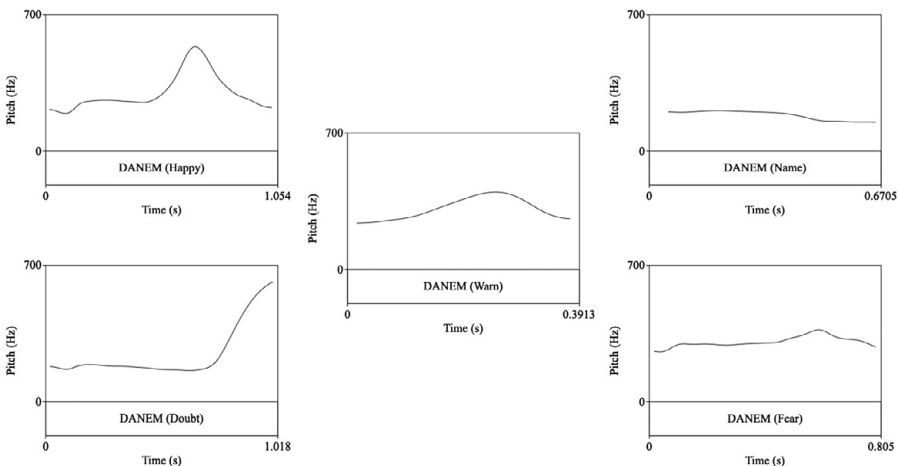

Figure 1. A sample noun (*danem*) with pitch tracks for the five emotion categories.

significant effect of emotional prosody category, $F(4, 60) = 82.24$, $p < .001$. The post hoc pairwise comparisons revealed that the mean f0 of all pairs of emotional prosody categories were significantly different from each other ($p < .01$), except between Warn and Fear. Overall, Name had the lowest mean f0 relative to the other emotional prosody categories. For the f0 range, there was a significant effect of emotional prosody category, $F(4, 60) = 17.68$, $p < .001$. The post hoc pairwise comparisons revealed that the mean f0 of all pairs of emotional prosody categories were significantly different from each other ($p < .01$), except between Name and Warn, Warn and Fear and Doubt and Happy. For the mean intensity, there was a significant effect of emotional prosody category, $F(4, 60) = 4.69$, $p = .002$. The post hoc pairwise comparisons revealed that the mean intensity between Name and Happy ($t(60) = -2.12$, $p = .04$); Doubt and Fear ($t(60) = -3.12$, $p = .003$); Doubt and Happy ($t(60) = -4.06$, $p < .001$) and Warn and Happy ($t(60) = -2.26$, $p = .03$) were significantly different. None of the other pairs were significantly different from each other. For maximum intensity, there was a significant effect of emotional prosody category, $F(4, 60) = 24.71$, $p < .001$. The post hoc pairwise comparisons revealed that the maximum intensity of all pairs of emotional prosody categories were significantly different from each other ($p < .01$), except between Name and Doubt and Warn and Happy. Both the pitch contour patterns and the acoustic measures of prosodic features show that the emotion categories are acoustically distinct from each other. Full details, including graphs of each dimension, are available on the OSF webpage.

2.1.2.3. Trained nouns and referents. For Study 1, the stimuli consisted of these 15 novel nouns paired with 15 novel pictures of either Aliens (from Gupta et al., 2004) or Objects (from Horst, 2016). These were selected from the same picture sets used by West et al. (2017, 2022). Training sets of the 15 noun-referent pairings were created with 3 words spoken from each of the 5 prosody types. There were five randomly assigned counterbalancing conditions (A, B, C, D, E) such that each word was presented in each prosody across participants in a partial Latin square design; for example, *tebos* was trained in a happy prosody in Condition A and trained in a fearful prosody in Condition B. Different versions of the training videos based on this counterbalancing variable ensured that the specific prosody (e.g., warning) was being tested across participants with different novel words and referents.

2.1.2.4. Training videos. Training in each block consisted of a video that presented each word with its corresponding referent, either an object or an alien, based on their between-participants referent condition. In the video, each trained referent was presented in isolation for 5 seconds, and the novel label was played with the assigned prosody at 1 second. The prosody for each noun-referent pairing was consistent throughout all training videos for each participant; for example, if participants heard the word *tebos* in a happy prosody in Block 1, they always heard *tebos* in a happy prosody during training. Each word was presented twice per training video in a predetermined random order to create a 2.5-min training video. This predetermined order of the stimuli in the training videos was identical in Blocks 1 and 3, and a different predetermined random order was used in Blocks 2 and 4.

2.1.2.5. Test blocks. Test trials in each block included a four-item test set for participants to choose the correct referent (A, B, C, or D). The position of the correct

answer was counterbalanced across trials for an approximately equal number of correct answers in each position. For Blocks 1–4, test sets included the trained item (correct answer), a foil item that was trained with a different word but the same prosody, a foil item trained on a different word and a different prosody, and a novel untrained item. Test sets were identical in Blocks 1–4, and test items were randomly presented in each block.

2.1.2.6. Generalization test block. A generalization test block included a subset of five of the trained items, which included the same word-referent pair for all participants. However, each item had been trained using a different prosody for different participants; thus, the generalization includes the five emotions, balanced across participants. The generalization four-item test set included the same three types of foil items as before, but the target item (correct answer) was not the trained item, but one from the same category requiring participants to generalize the trained word to a new exemplar of the referent. The new alien exemplars differed in aspects such as slight differences in head shape, clothing patterns or clothing configurations and slight differences in body shape. The new object exemplars differed in color and/or positioning of the object parts. For example, the correct answer in Figure 2 is the blue alien, but the blue alien in the generalization set is wider and has taller boots. Similarly, the elongated object is the correct answer, but the one pictured in the generalization set is yellow with a green end instead of orange with a blue end (see Figure 2 for sample trained items and resulting test sets).

2.1.2.7. Catch trials. Participants also received catch trials presenting an entirely new untrained set of four items in their test set for which they were told, 'Please select A' with approximately equal instructions to select A, B, C, and D across the entire study. Each test block included 15 test trials and 2 catch trials, and the Generalization Block included 5 test trials (subset of 5 of the trained items) and 2 catch trials.

2.1.3. Design and procedure

The overall design for this study was 2 between-participants Referent Animacy (Alien, Object) × 5 within-participants Trained Prosody (Doubt, Fear, Happy, Name, Warn) × 5 between-participants counterbalancing variable Training Sets (A, B, C, D, E) × 5 within-participants Test Blocks (1, 2, 3, 4, Generalization).

The study was presented online through Qualtrics. After consenting to participate, the participants were instructed to adjust their speakers to a comfortable volume. Then, they were informed that they would watch a video demonstrating new words that go with pictures, and that their job was to learn these words. They were told they would see a video and be tested several times, and they were to see if they could get better each time. At the end of the study, the participants answered demographic questions and were debriefed.

Participants watched a 2.5-min raining video to learn 15 novel noun-referent pairings in a predetermined randomized presentation order. The participants completed a recognition test in which they heard the word in a neutral tone by the same female speaker from the training (e.g., which one is a *balide*?) and were asked to select the correct referent when paired with three foil items. Participants were required to answer to move to the next test item, and they did not receive feedback. This train–test sequence was repeated four times in total. Then, the participants completed a generalization task for five of the trained items. All participants received the same five
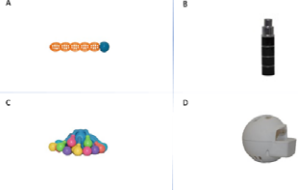
Trained Item	Test Set	
Balide		
Balide		
	Generalization Set	
Balide		
Balide		

Figure 2. An example novel noun label paired with a trained item for the Alien and Object Conditions with each item's corresponding Test Set and Generalization Set. The correct answer is A for each of these examples.

items, although their trained prosody differed based on their Training Set condition, and the Generalization Block seamlessly began without further training.

2.2. Results

Test Block 1 accuracy scores were above 25% chance performance across all five prosodies, with an Exact Binomial 95% CI for the lowest accuracy being 43.16%–

52.28%, p's < .001, indicating that participants had an initial understanding of the word even by their first test.

The accuracy scores were further analyzed using a logistic mixed model with the predictor variables of Referent Condition (Object, Alien); Prosody (Doubt, Fear, Happy, Name, Warn) and Test Block as a numeric variable (1, 2, 3, 4, 5). Random effects were included for each participant and each word to account for the dependence within each participant's responses and the inherent tendencies associated with each nonsense word used in the study. An analysis of deviance was used to compare the main effects only model against the interaction model, favoring the main effects model, with a nonsignificant effect of the interaction model. The Bayesian information criterion also favors the main effects model (BIC = 1.0405×10^4) over the interaction model (BIC = 1.0436×10^4), with a lower BIC indicating a more favorable model.

Are there differences across emotional prosody conditions in word learning? Using an analysis of deviance methodology, with Name as the reference level, Prosody significantly predicted accuracy X^2 (4, $N = 160$) = 21.86, $p < .001$, with Warn prosody less accurate than Name, ($b = -0.21$, $z = -2.68$, $p < .01$). How does learning progress over time? Test Block also significantly predicted accuracy, X^2 (1, $N = 160$) = 771.88, $p < .001$. Backward difference contrasts, considering each level against its previous level, indicate improvement from Block 1 to Block 2 ($b = 1.18$, $z = 16.77$, $p < .001$), Block 2 to Block 3 ($b = 0.48$, $z = 6.52$, $p < .001$) and Block 3 to Block 4 ($b = 0.17$, $z = 2.14$, $p < .05$), demonstrating decreasing improvement with each block. This analysis also indicates a significant decrease from Block 4 to the Generalization Block ($b = -0.26$, $z = -2.32$, $p < .05$), as would be expected with the more difficult task of generalizing the word to a novel exemplar. Does the type of referent matter? The Referent Condition also significantly predicted accuracy, with Objects identified more accurately than Aliens, $b = 0.76$, $z = 3.67$, $p < .001$ (see Figure 3 for means and standard errors).
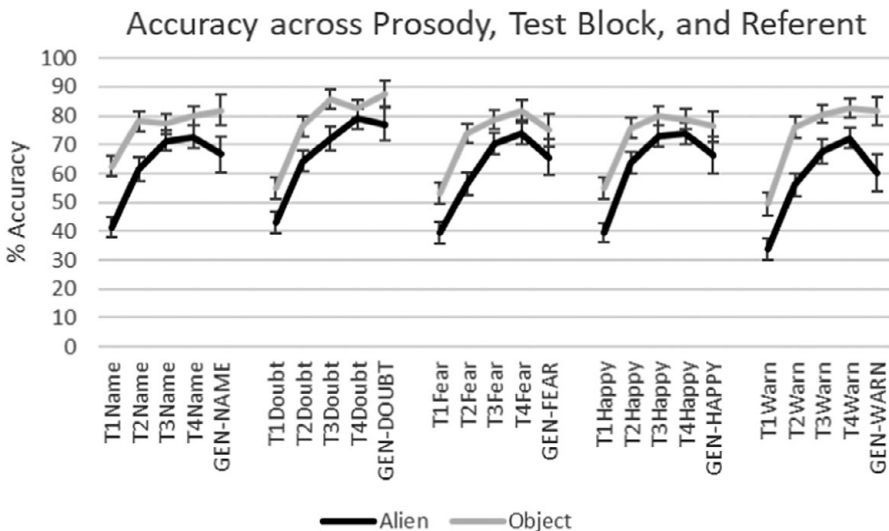

Figure 3. Graph represents accuracy scores for each Test Block across each Prosody Type and for each Referent Condition. Bars represent standard error of the mean.

2.3. Discussion

The results of Study 1 show differences across prosody and across referents in word learning, with Aliens worse than Objects and Warn as less accurate than Name. West et al. (2017) showed better learning in neutral over fearful conditions (and in neutral over happy conditions for those with low autism scores), suggesting that emotion hurts memory for labeling the alien stimuli. The current study shows worse learning with the Warn category than with the Name category. The distinctions in word learning across emotion prosody type suggests that emotion may have a nuanced effect on word learning, with some types of emotions as detrimental to a word learning task. The West et al. (2022) study with children showed no memory distinction for learning novel object labels by prosody. In the current study, adults showed an effect of prosody with both types of referents.

There were also differences in word learning by referent, with Objects learned better than Aliens. The labels were identical in each of these conditions; therefore, the difference must be due to the images themselves. The Aliens appeared to be animate creatures with complex coloring, limbs and clothing. The Objects were simple and toy-like, with fewer parts than the Aliens. The difference in memory for Objects over Aliens is surprising because there is evidence demonstrating better learning of animate words, both in printed words (Nairne et al., 2013; Popp & Serra, 2016) and in pictures (Bonin et al., 2014). It may be the complexity of the referents and not the animacy that differentiates word learning, as simple objects are remembered better than complex objects (Eng et al., 2005; Mishra, 1984). To test the effects of referent animacy on novel noun learning while controlling for referent complexity, in Study 2, Objects with Faces were created to contrast with Objects.

3. Study 2: Objects versus objects with faces

The goal of Study 2 is to examine the role of animacy in adults' ability to learn novel labels for animate and inanimate objects across five prosodic conditions. Are there differences in word learning across prosodic conditions? Does the animacy of the referent matter? How does learning progress over time? It is anticipated that the prosodic condition will influence word learning, which would replicate Study 1. In addition, there will be improved word learning across time.

3.1. Method

3.1.1. Participants

The participants included 237 undergraduate students from a regional campus of a large university who participated for research credit in their introductory psychology course. None of the participants participated in Study 1. Participants were excluded from the analyses for a variety of reasons: failure to complete the entire session ($n = 44$), non-native English speakers ($n = 29$), uncorrected hearing or vision problems ($n = 13$) or below 80% on catch trials ($n = 23$), for a final sample of 128 (*M* age = 19.41 years, *SD* = 4.65). There were 50 males (39%), 74 females (58%), 3 nonbinary, gender-fluid or agender (2%) and 1 undisclosed (1%). The Institutional Review Board of the authors' home institution granted ethics approval for this project, and the participants provided informed consent prior to participation.

3.1.2. Materials

The same 15 words, across the 5 prosodies from Study 1 were used in Study 2. The stimuli consisted of 15 novel nouns paired with 15 novel pictures of either Objects or Objects with Faces. The Objects came from Study 1, and a smile and eyes were added to each to create Objects with Faces (see Figure 4 for sample trained items and resulting Test Sets). Just as in Study 1, training consisted of a video that presented each word and its corresponding referent, either an Object or an Object with Face based on their
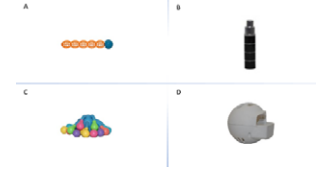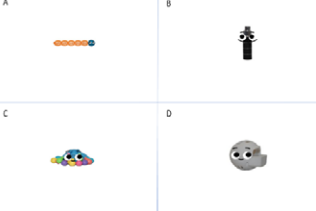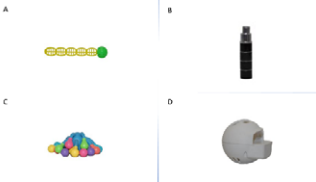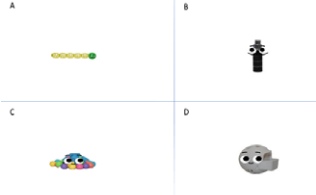

Figure 4. An example novel noun label paired with a trained item for the Object and Object with Faces Conditions with each item's corresponding Test Set and Generalization Set. The correct answer is A for each of these examples.

between-participants referent condition. The training videos and counterbalancing measures were identical to Study 1 but using Objects with Faces instead of Aliens. The Test Block, Generalization Block and Catch Trials were the same as those in Study 1.

3.1.3. Design and procedure

The overall design for this study was 2 between-participants Referent Animacy (Object, Object with Faces) × 5 within-participants Trained Prosody (Doubt, Fear, Happy, Name, Warn) × 5 between-participants counterbalancing variable Training Sets (A, B, C, D, E) × 5 within-participants Test Blocks (1, 2, 3, 4, Generalization). The procedure was identical to that in Study 1.

3.2. Results

Test Block 1 accuracy scores were above 25% chance performance across all five prosodies, with an Exact Binomial 95% CI for the lowest accuracy being 49.74%–54.86%, p's < .001, indicating that participants had an initial understanding of the word even by their first test.

The accuracy scores were further analyzed using a logistic mixed model with the predictor variables of Referent Condition (Object, Object with Faces), Prosody (Doubt, Fear, Happy, Name, Warn) and Test Block as a numeric variable (1, 2, 3, 4, 5). Random effects were included for each participant and each word to account for the dependence within each participant's responses and the inherent tendencies associated with each nonsense word used in the study. An analysis of deviance was used to compare the main effects only model against the interaction model, favoring the main effects model, with a nonsignificant effect of the interaction model. The Bayesian information criterion also favors the main effects model (BIC = 8242.41) over the interaction model (BIC = 8276.58), with a lower BIC indicating a more favorable model.

Are there differences across emotional prosody conditions in word learning? Using an analysis of deviance methodology, with Name as the reference level, Prosody significantly predicted accuracy X^2 (4, N = 128) = 22.18, p < .001, with Warn prosody once again less accurate than Name, ($b = -0.28, z = -3.16, p < .01$). How does learning progress over time? The Test Block also significantly predicted accuracy, X^2 (1, N = 128) = 680.23, p < .001. Backward difference contrasts, considering each level against its previous level, indicate improvement from Block 1 to Block 2 ($b = 1.15$, $z = 14.64, p < .001$), Block 2 to Block 3 ($b = 0.57, z = 6.93, p < .001$) and Block 3 to Block 4 ($b = 0.21, z = 2.44, p < .05$), demonstrating decreasing improvement with each block. This analysis also indicates no significant change from Block 4 to the Generalization Block ($b = -0.11, z = -0.84, p > .05$), as would be expected without additional training. Does the animacy of the referent matter? Referent Condition did not predict accuracy, with Objects identified at the same level as Faces, $b = -0.08, z = -0.37, p > .05$ (see Figure 5 for the means and standard errors).

3.3. Discussion

There were word learning differences by prosodic category in Study 2, with Warn learned less accurately than Name. There was also an improvement in word learning across time for all prosody categories. Unlike Study 1, there were no differences in performance between the two referent categories (Objects, Objects with Faces). The
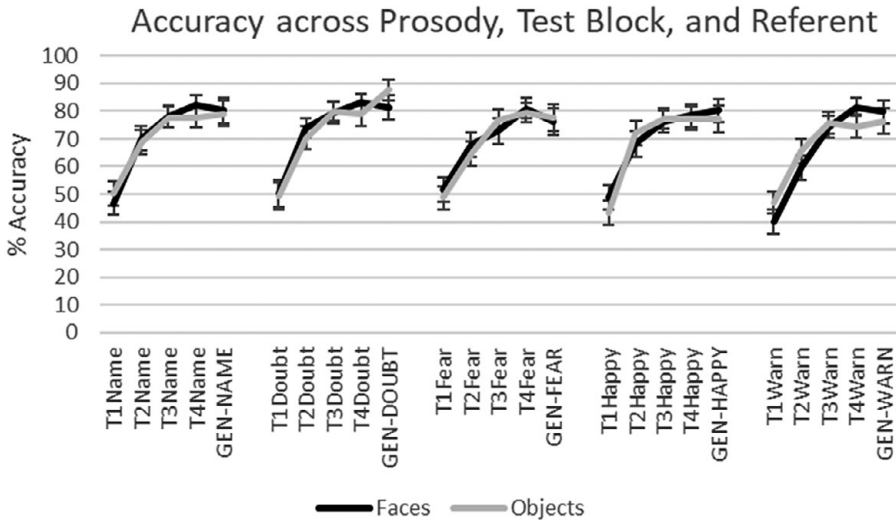
Figure 5. Graph represents accuracy scores for each Test Block across each Prosody Type and for each Referent Condition. Bars represent standard error of the mean.

addition of eyes and a mouth made objects appear animate, but this did not translate into a difference in word learning. Both Objects and Objects with Faces were similar in complexity, with the only difference being the addition of a face. Study 2 demonstrated similar effects of prosody on word learning across simple animate and inanimate objects.

4. General discussion

Word learning improved with training for all prosody categories, and all training and stimuli conditions led to successful word generalization. The current set of studies showed less accurate memory performance for Warn than for Name in both Studies 1 and 2. The result of differences by emotional prosody is similar to previous research showing a fear condition interfering with word-learning relative to a neutral condition for typical adults for Alien stimuli (West et al., 2017). Study 2 found differences in word learning by prosody for Objects, which differs from the previous work by West et al. (2022). However, West et al. (2022) tested novel word learning for Objects with 7- to 9-year-olds and found no differences in learning. The current research demonstrates that emotional prosody similarly affected all stimuli for adults, including Aliens, Objects and Objects with Faces.

There is evidence that animate words are recalled better than inanimate words (Aka et al., 2020; Bonin et al., 2014; Meinhardt et al., 2019; Nairne et al., 2013; Popp & Serra, 2016). However, these studies present words in printed form or in lists. Animacy did matter in Study 1, but inanimate objects were remembered better than animate objects. The memory difference between Objects and Aliens may relate to factors other than animacy. The Aliens were complex images, and differences between them included color, clothing, height, number of arms, tail and so forth. The Objects, on the other hand, were relatively simple images that differed by shape

and number of parts and were made to resemble children's toys. Perhaps the reason the Objects were better learned was because they were more distinctive and easily distinguished from each other (Eng et al., 2005; Mishra, 1984).

Why does emotional prosody influence learning novel nouns? One explanation for the results is attention. Successful word learning requires attention to and encoding both auditory and visual information and then integrating that information (Bhat et al., 2021; Bruce et al., 2022; Samuelson et al., 2017; Smith et al., 2010). Salient auditory information can capture attention with an orienting response (SanMiguel et al., 2009). Pulling attention to one modality reduces attention and therefore processing in another modality. For example, salient information in one modality (e.g., auditory) draws attention to that modality and thereby disrupts processing in another (e.g., visual) modality (Cowan & Barron, 1987; Elliott et al., 2014; Francis et al., 2017; Robinson et al., 2016). Disproportionately salient auditory information may disrupt encoding visual attention or interfere with the binding of the visual and auditory information that is necessary for successful word learning (Bhat et al., 2021).

The various emotional prosodies differed in their ability to grab the auditory attention. The Warn category is most likely to cause an orienting response because it suggests danger. The Warn stimuli were shortest on average, but they also had the highest maximum intensity and one of the highest mean f0 values. Not surprisingly, Warn was the category that was learned less accurately than Name in both studies. A similar result is seen in the Fear category of West et al. (2017), which could also be argued to pull attention. Attention can be conceptualized as a limited resource for storage and processing that is linked to working memory (Oberauer, 2019). The emotional prosody was not needed to learn the novel noun labels; however, the emotion could serve as a distraction. There was no difference in noun memory for children by emotion category, but West et al. (2022) found that children spent less time looking at the objects themselves in their fear condition, which also happened to be the worst word learning condition in adults. West et al. (2017) similarly argue that the emotional information may have interfered with attentional allocation.

It is also possible for emotion to enhance memory. There is much evidence that emotional events are better remembered than neutral events (Tyng et al., 2017). Emotionally charged stimuli, particularly those that are perceived as threats, are selected through attention via the amygdala (Vuilleumier, 2005), which may work alongside the hippocampal complex to focus and create episodic memories of emotional events (Phelps, 2004). Emotional events are remembered better; however, there are distinctions between memory for details and concepts depending on the valence of the stimuli (Kensinger, 2009). The memory-enhancing effect for emotional events may operate differently than the appropriate attentional allocation required for word-learning events. Adults differentially learned novel nouns across multiple prosodic intentions, similar to the findings of West et al. (2017). This result is particularly striking because prosody was irrelevant to the word learning task. We have suggested attention as a possible mechanism to explain these differences in word learning. Future studies could more directly measure attention using tools such as eye tracking. Future studies could also test separate acoustic features (f0, intensity etc.) to determine which aspects of speech might influence word learning and distinguish between the emotional prosody conveyed by an acoustic pattern and the specific acoustic dimensions of the stimuli.

West et al. (2022) claim that their differential findings across experiments could be explained by developmental mechanisms, which may be influencing the effects of

prosody on word learning. This is still a realistic possibility given the prosodic development that takes place during childhood (Armstrong & Hübscher, 2018; Graham et al., 2017; Hupp & Jungers, 2009) and later adulthood (Sober et al., 2016). For example, positive prosody has been shown to enhance learning content over neutral prosody for 11- to 13-year-old children, whereas it had no effect on 8- to 10-year-old children (Dylman & Champous, 2022). Adults and children could be differentially attending to the prosodic information, and this would be worth investigating with preschoolers and older adults to determine how emotional prosody affects word learning. We also propose an attentional account for the differential findings across prosodies, which may affect children and adult word learners differently. Children have less attentional control than adults (Posner & Rothbart, 2007; Ruff & Capozzoli, 2003) to ignore irrelevant prosodic information. We know that improved attentional mechanisms are related to word learning (Smith et al., 2010), and we know that children have potentially more reliance on prosody than adults (de Diego-Balaguer et al., 2016). Therefore, additional research is needed to investigate how various irrelevant prosodic labels affect young children's noun learning. Given the artificial nature of the current word learning task and the oversimplification of animacy used in this research, future research should also investigate the role of prosody on noun learning in a more ecologically valid way with words or phrases extracted from real speech exhibiting various prosodic speech acts as well as a more lifelike representation of animacy.

Adults can learn novel nouns across a variety of prosodic labels; however, the labels that had the potential to draw attention from the task with strong prosody (Warn!) are the labels that appear to be learned less accurately compared to naming. Future research should further investigate the specific effects of attention to prosody, how this affects word learning across grammatical classes and across languages, and how this process develops in children. Understanding how individuals learn new words with a variety of prosodies has the potential to impact how we teach languages, instruct children and present new information or terminology to the public.

Competing interest. The authors declared none.

References

Aka, A., Phan, T. D., & Kahana, M. J. (2020). Predicting recall of words and lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47, 765–784. https://doi.org/10.1037/xlm0000964

Armstrong, M. E., & Hübscher, I. (2018). Children's development of internal state prosody. In P. Prieto, & N. Esteve-Gibert (Eds.), *The development of prosody in first language acquisition* (pp. 271–293). John Benjamins Publishing Company.

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, F., & van Lieshout, P. H. H. M. (2016). Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test for rating of emotions in speech. *Journal of Speech, Language, and Hearing Research*, 59, 72–89. https://doi.org/10.1044/2015_JSLHR-H-14-0323

Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology*, 107, 87–99. https://doi.org/10.1016/j.jecp.2010.04.012

Berman, J. M. J., Graham, S. A., Callaway, D., & Chambers, C. G. (2013). Preschoolers use emotion in speech to learn new words. *Child Development*, 84, 1791–1805. https://doi.org/10.1111/cdev.12074

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2021). Word-object learning via visual exploration in spaces (WOLVES): A neural process model of cross-situation word learning. *Psychological Review*, 129, 640–695. https://doi.org/10.1037/rev0000313

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cognition*, 42, 370–382. https://doi.org/10.3758/s13421-013-0368-8

Bruce, M., Panneton, R., & Taylor, C. (2022). Multisensory integration and maternal sensitivity are related to each other and predictive of expressive vocabulary in 24-month-olds. *Journal of Experimental Child Psychology*, 214, 105304. https://doi.org/10.1016/j.jecp.2021.105304

Cowan, N., & Barron, A. (1987). Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Perception & Psychophysics*, 41, 393–401. https://doi.org/10.3758/bf03203031

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201. https://doi.org/10.1177/002383099704000203

de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology*, 7, 44. https://doi.org/10.3389/fpsyg.2016.00044

Dylman, A. S., & Champous, L. M. (2022). The effect of emotional prosody on content learning in Swedish school children. *Applied Cognitive Psychology*, 36, 1339–1345. https://doi.org/10.1002/acp.4009

Elliott, E. M., Morey, C. C., Morey, R. D., Eaves, S. D., Tally Shelton, J. & Lutfi-Proctor, D. A. (2014). The role of modality: Auditory and visual distractors in Stroop interference. *Journal of Cognitive Psychology*, 26, 15–26. https://doi.org/10.1080/20445911.2013.859133

Eng, H. Y., Chen, D. & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, 12, 1127–1133. https://doi.org/10.3758/BF03206454

Francis, W. S., MacLeod, C. M., & Taylor, R. S. (2017). Joint influence of visual and auditory words in the Stroop Task. *Attention, Perception & Psychophysics*, 79, 200–211. https://doi.org/10.3758/s13414-016-1218-0

Graham, S. A., San Juan, V., & Khu, M. (2017). Words are not enough: How preschoolers' integration of perspective and emotion informs their referential understanding. *Journal of Child Language*, 44, 500–526. https://doi.org/10.1017/S0305000916000519

Gupta, P., Lipinski, J., Abbs, B., Lin, P.-H., Aktunc, E., Ludden, D., Martin, N., & Newman, R. (2004). Space Aliens and nonwords: Stimuli for investigating the learning of novel word-meaning pairs. *Behaviour Research Methods, Instruments, & Computers*, 36, 599–603. https://doi.org/10.3758/BF03206540

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, 38, 70–86. https://doi.org/10.1016/j.jml.2016.01.001

Horst, J. S. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. Databrary. Retrieved October 8, 2021 from https://doi.org/10.17910/B7.209.

Hupp, J. M., & Jungers, M. K. (2009). Speech priming: An examination of rate and syntactic persistence in preschoolers. *British Journal of Developmental Psychology*, 27, 495–504. https://doi.org/10.1348/026151008X345988

Hupp, J. M., & Jungers, M. K. (2013). Beyond words: Comprehension and production of pragmatic prosody in adults and children. *Journal of Experimental Child Psychology*, 115, 536–551. https://doi.org/10.1016/j.jecp.2012.12.012

Hupp, J. M., Jungers, M. K., Hinerman, C. M., & Porter, B. P. (2021). Cup! Cup? Cup: Comprehension of intentional prosody in adults and children. *Cognitive Development*, 57, 100971. https://doi.org/10.1016/j.cogdev.2020.100971

Ito, K., & Speer, S. R. (2008). Anticipatory effect of intonation: Eye movements during instructed visual search. Journal of Memory and *Language*, 58, 541–573. https://doi.org/10.1016/j.jml.2007.06.013

Kensinger E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review: Journal of the International Society for Research on Emotion*, 1, 99–113. https://doi.org/10.1177/1754073908100432

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, 31, 1169–1180. https://doi.org/10.3758/BF03195800

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112, 473–481. https://doi.org/10.1016/j.cognition.2009.06.007

Krestar, M. L., & McLennan, C. T. (2019). Responses to semantically neutral words in varying emotional intonations. *Journal of Speech, Language, and Hearing Research*, 62(3), 733–744. https://doi.org/10.1044/2018_JSLHR-H-17-0428

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7, 106–122. https://doi.org/10.1006/jpho.1997.0046

Lehiste, I., Olive, J. P., & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustic Society of America*, 60, 1199–1202. https://doi.org/10.1121/1.381180

Lei, Z., Bi, R., Mo, L., Yu, W., & Zhang, D. (2021). The brain mechanism of explicit and implicit processing of emotional prosodies: An fNIRS study. *Acta Psychologica Sinica*, 53, 15–25. https://doi.org/10.3724/SP.J.1041.2021.00015

Lieberman, P., & Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *Journal of the Acoustic Society of America*, 34, 922–927. https://doi.org/10.1121/1.1918222

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception & Performance*, 13, 384–394. https://doi.org/10.1037//0096-1523.13.3.384

Mauchand, M., & Pell, M. D. (2021). Emotivity in the voice: Prosodic, lexical, and cultural appraisal of complaining speech. *Frontiers in Psychology*, 11. https://doi.org/10.3389/fpsyg.2020.619222

Meinhardt, M. J., Bell, R., Buchner, A., & Röer, J. P. (2019). Adaptive memory: Is the animacy effect on memory due to richness of encoding? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46, 416–426. https://doi.org/10.1037/xlm0000733

Mishra, R. C. (1984). Recognition memory for visual stimuli as a function of labeling. *Psychological Studies*, 29, 4–8.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93, 1097–1108. https://doi.org/10.1121/1.405558

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science*, 24, 2099–2105. https://doi.org/10.1177/0956797613480803

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, 33, 127–146. https://doi.org/10.1111/j.1551-6709.2008.01007.x

Oberauer, K. (2019). Working memory and attention: A conceptual analysis and review. *Journal of Cognition*, 2, 1–23. https://doi.org/10.5334/joc.58

Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, 14, 198–202. https://doi.org/10.1016/j.conb.2004.03.015

Popp, E. Y., & Serra, M. J. (2016). Adaptive memory: Animacy enhances free recall but impairs cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 186–201. https://doi.org/10.1037/xlm0000174

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, 58, 1–23. https://doi.org/10.1146/annurev.psych.58.110405.085516

Reinisch, E., Jesse, A. & Nygaard, L. C. (2013). Tone of voice guides word learning in informative referential contexts. *The Quarterly Journal of Experimental Psychology*, 66, 1227–1240. https://doi.org/10.1080/17470218.2012.736525

Robinson, C. W., Chandra, M., & Sinnett, S. (2016). Existence of competing modality dominances. *Attention, Perception, and Psychophysics*, 78, 1104–1114. https://doi.org/10.3758/s13414-016-1061-3

Ruff, H. A., & Capozzoli, M. C. (2003). Development of attention and distractibility in the first 4 years of life. *Developmental Psychology*, 39(5), 877–890. https://doi.org/10.1037/0012-1649.39.5.877

Samuelson, L. K., Kucker, S. C., & Spencer, J. P. (2017). Moving word learning to a novel space: A dynamic systems view of referent selection and retention. *Cognitive Science*, 41, 52–72. https://doi.org/10.1111/cogs.12369

SanMiguel, I., Linden, D., & Escera, C. (2009). Attention capture by novel sounds: Distraction versus facilitation. *European Journal of Cognitive Psychology*, 22, 481–515. https://doi.org/10.1080/09541440902930994

Shintel, H., Anderson, N. L., & Fenn, K. M. (2014). Talk this way: The effect of prosodically conveyed semantic information on memory for novel words. *Journal of Experimental Psychology: General*, 143, 1437–1442. https://doi.org/10.1037/a0036605

Shintel, H., & Nusbaum, H. C. (2007). The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. *Cognition*, 105, 681–690. https://doi.org/10.1016/j.cognition.2006.11.005

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, 55, 167–177. https://doi.org/10.1016/j.jml.2006.03.002

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually cued attention and early word learning. *Cognitive Science*, 34, 1287–1314. https://doi.org/10.1111/j.1551-6709.2010.01130.x

Snefjella, B., Lana, N., & Kuperman, V. (2020). How emotion is learned: Semantic learning of novel words in emotional contexts. *Journal of Memory and Language*, 115, 104171. https://doi.org/10.1016/j.jml.2020.104171

Sober, J. D., VanWormer, L. A., & Arruda, J. E. (2016). Age-related differences in recall for words using semantics and prosody. *The Journal of General Psychology*, 143, 67–77. https://doi.org/10.1080/00221309.2015.1073138

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, 8, 1454. https://doi.org/10.3389/fpsyg.2017.01454

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Science*, 9, 585–594. https://doi.org/10.1016/j.tics.2005.10.011

Warren, P. (1997). Prosody and language processing. In: S. Garrod & M. Pickering (Eds.), *Language processing* (pp. 155–188). Psychology Press.

West, M. J., Angwin, A. J., Copland, D. A., Arnott, W. I., & Nelson, N. L. (2022). Effects of emotional cues on novel word learning in typically developing children in relation to broader autism trait. *Journal of Child Language*, 49, 503–521. https://doi.org/10.1017/S0305000921000192

West, M. J., Copland, D. A., Arnott, W. L., & Nelson, N. L. (2017). Effects of emotional prosody on novel word learning in relation to autism-like traits. *Motivation & Emotion*, 41, 749–759. https://doi.org/10.1007/s11031-017-9642-6