Development of sensitivity to beat gesture and contrastive accenting in support of word learning in early childhood in boys and girls

Laura M. Morett, Cailee M. Nelson, Sarah S. Hughes-Berheim and Jason Scofield
The University of Alabama, USA

Abstract
This research investigated whether observing beat gesture and hearing contrastive accenting with novel words enhances their learning in early childhood and whether these effects differ by sex in light of sex differences in the pace of language development. Fifty-three 3- to 5-year-old boys and girls learned pairs of novel words with contrasting referents with beat gesture, contrastive accenting, both, or neither. Knowledge of these words was then tested via a referent identification task. Novel word learning did not differ by beat gesture or contrastive accenting, nor did use of these cues to support word learning differ by age. However, 3-year-old boys were better able to identify the referents of novel words learned with rather than without beat gesture, and boys’ ability to identify the referents of novel words learned without beat gesture improved from ages 3 to 5 years. By contrast, no such effects of beat gesture on novel word learning by age were observed for girls. These results suggest that, for 3-year-old boys, beat gesture may compensate for difficulty deducing contrast from speech alone, and that their reliance on beat gesture as a cue to contrast decreases as their ability to deduce contrast from speech improves during early childhood. Thus, beat gesture may serve as a visual cue to contrast that scaffolds young children’s learning of words with contrasting meanings by supplementing the use of cues to contrast conveyed via speech.

Keywords
Beat gesture, contrastive accenting, word learning, early childhood, sex differences

Corresponding author:
Laura M. Morett, Department of Educational Studies in Psychology, Research Methodology, and Counseling, The University of Alabama, Box 870240, Tuscaloosa, AL 35487, USA.
Email: lmorett@ua.edu
In light of the amount of information conveyed via language, it is crucial for children to determine what is most important so they can concentrate on it. Focus, which consists of new, non-derivable, or contrastive information conveyed via a word or phrase, can be expressed via several cues (Lambrecht, 1996), all of which children must learn to interpret and produce language effectively (Dimroth & Narasimhan, 2012). Two of these cues, which often occur conjointly, are pitch accenting, consisting of changes in acoustic characteristics of speech such as frequency, intensity, and duration (Ladd, 1996); and beat gesture, consisting of non-referential, rhythmic movements (McNeill, 2005). When pitch accenting and beat gesture consistently co-occur with contrasting alternatives in discourse (e.g. THIS vs that), they serve as reliable cues to contrast, enhancing attention to and memory for them in adults (Fraundorf et al., 2010; Kushch & Prieto, 2016; Morett et al., 2021; Morett & Fraundorf, 2019; Morett, Roche, et al., 2020) as well as children (Igualada et al., 2017; Ito et al., 2012, 2014; Lee & Snedeker, 2016; Llanes-Coromina et al., 2018; Sekerina & Trueswell, 2012). At present, however, it is uncertain whether pitch accenting and beat gesture can facilitate acquisition of contrasting sets of novel words differing in meaning, so their utility in the context of word learning is unclear. The current research investigates interpretation of contrastive accenting and beat gesture as cues to contrast in the context of word learning during early childhood, illuminating whether these cues enhance young children’s ability to learn and differentiate between contrasting pairs of novel words.

**Pitch accenting**

Pitch accenting refers to a phonological construct conveying prominence that is realized acoustically. In English, pitch accenting is realized on the word or phrasal level and indicates focus when used contrastively (e.g. red ball [a ball that happens to be red] vs RED ball [a ball that is red, in contrast to the color of another ball]); by contrast, stress is realized on the segmental level and indicates word class when used contrastively (e.g. PREsent [noun] vs preSENT [verb]). The acoustic correlates of both cues in English include fundamental frequency ($F_0$), duration, and intensity and may differ in relation to one another (e.g. pitch accenting may be conveyed primarily via variation in $F_0$, whereas stress may be conveyed primarily via variation in intensity; Bolinger, 1958). By contrast, in pitch accent languages such as Japanese, pitch accenting is realized solely via variation in $F_0$ on the segmental level and indicates a difference in meaning when used contrastively (e.g. 雨: [high-low; rain] vs 餅: [low-high; candy]). In English, two of the most common pitch accents are presentational accenting, which consists of a target pitch with $F_0$ high in the talker’s range, and contrastive accenting, which consists of an initial low pitch followed by a sharp rise to a high target pitch on the accented syllable (Pierrehumbert & Hirschberg, 1990). Presentational accenting is used to convey novelty of information in general, whereas contrastive accenting is used specifically to convey contrast with other parts of a discourse. An example of a discourse with both presentational and contrastive pitch accenting follows:

1a. [S1] What did John buy at the store?
1b. [S2] He bought some shirts.
2a. [S1] Did you say he bought some shorts?
2b. [S2] No, I said he bought some **SHIRTS**.

In the above discourse, the word ‘shirts’ is presented as new information in (1b). However, in (2b), the word ‘shirts’ contrasts with another referent (i.e. shorts), and would therefore likely be contrastively accented. Although this example involves correction, contrastive accenting can be used to convey contrast without correction, as in the following discourse, which can stand alone:

3a. John bought **SHIRTS**, not shorts.

These differences in the informational contributions of presentational and contrastive pitch accenting are evident in memory for and processing of spoken discourse. In English-speaking adults as well as children, memory for referents with contrastive accenting is superior to memory for referents with presentational accenting, particularly when a salient contrasting referent must be rejected (Fraundorf et al., 2010, 2012; Lee & Snedeker, 2016; Sanford et al., 2006). Moreover, in English-, Japanese-, and German-speaking adults and children, contrastive accenting directs attention to contrasts between referents (Ito et al., 2012; Ito & Speer, 2008; Kurumada et al., 2014; Watson et al., 2008; Weber et al., 2006). Thus, contrastive accenting helps both adults and children to attend to and remember specific alternatives contrasting with mentioned and unmentioned alternatives.

In general, evidence from research suggests that comprehension of contrastive pitch accenting develops during early childhood (Ito et al., 2012, 2014; Sekerina & Trueswell, 2012; Wells et al., 2004). In 5- to 6-year-old Japanese-speaking children, eye gaze patterns demonstrate anticipation of a color contrast when a prenominal color word is contrastively accented (e.g. ‘Find the red cat. Now, find the **blue**. . .’ Ito et al., 2012). Compared with adults, however, 5–6-year-old speakers of Japanese as well as English and Russian are less efficient in processing contrastive pitch accenting and sometimes need additional cues to accurately recognize contrastive alternatives with it (Ito et al., 2012, 2014; Sekerina & Trueswell, 2012; Wells et al., 2004). Taken together, these findings suggest that comprehension of contrastive pitch accenting may not be fully mature even in school-aged children and that it continues to develop during early childhood, paralleling and possibly supporting the development of other attributes of language.

**Beat gesture**

To understand how prominence and contrast are conveyed visually via co-speech gestures (heretofore, gestures), it is necessary to understand the information conveyed via different types of gestures. According to McNeill (1992, 2005), co-speech gestures are meaningful hand movements produced in conjunction with speech consisting of two types: referential (i.e. iconic, metaphorical, deictic), which refer to a specific entity, and non-referential (i.e. beat), which do not. Beat gestures typically take the form of punctate hand movements (e.g. hand flicks) tightly linked to speech prosody that are used to convey prominence, serving as a cue to focus. Prototypically, beat gestures take the form of
downward hand flicks, but they can be articulated using other parts of the body (e.g. finger movements, head nods, foot taps), in other orientations (e.g. horizontal, oblique, curved), and with multiple components (Shattuck-Hufnagel & Ren, 2018). Notably, beat gestures are not mutually exclusive with referential gestures and can be combined with or superimposed over them (McNeill, 1992, 2005). Growing evidence has expanded upon McNeill’s claims about the form, prominence, and referentiality of beat gestures. In addition to hand flicks, beat gestures may take the form of an open palm facing outwards, a closed fist, a raised index finger, or a precision grip, all of which convey different epistemic effects (Lempert, 2011; Maricchiolo et al., 2009; Shattuck-Hufnagel et al., 2016; Streeck, 2008). Moreover, some beat gestures are not temporally aligned with pitch accenting (Rohrer et al., 2019; Shattuck-Hufnagel & Ren, 2018), and beat gestures can convey certain aspects of meaning in conjunction with concurrent speech (Alexanderson et al., 2013; Gluhareva & Prieto, 2017; Krahmer et al., 2002; Morett, Roche et al., 2020; Morett et al., 2021; Morett & Fraundorf, 2019; Yap et al., 2018). Critical to the purpose of the current work, when beat gesture consistently co-occurs with contrastive alternatives, it is interpreted as a cue to contrast in English as well as Catalan (Llanes-Coromina et al., 2018; Morett et al., 2021; Morett, Roche, et al., 2020). Thus, some contemporary researchers have proposed that all gestures, including beats, may have both prosodic and pragmatic components and can express a range of functions (Prieto et al., 2018; Shattuck-Hufnagel & Prieto, 2019).

Research examining development of beat gesture has investigated both production and comprehension. Although the developmental relationship between manual and vocal production is dynamic, manual production generally precedes vocal production in development due to greater motor control of the hands than the vocal articulators (Acredolo & Goodwyn, 1990). Thus, it is possible that beat gesture production may precede contrastive accenting production in development, as suggested by work indicating that development of beat gesture precedes prosody in focus production (Esteve-Gibert et al., 2022). Research investigating language development in bilingual children has shed light on the developmental trajectory of beat gesture production in childhood. French-English bilingual children produce mostly deictic gestures as they learn to speak and produce iconic and beat gestures more frequently as their speech increases in complexity. Between 3 and 6 years of age, however, these children still do not produce iconic and beat gestures with the same frequency as adults (Mayberry & Nicoladis, 2000; Nicoladis et al., 1999). Indeed, although beat gesture production increases during early childhood, it is immature at its culmination and continues to develop throughout middle childhood and adolescence in English- as well as French-speaking children (Colletta et al., 2010; Mathew et al., 2018).

Beat gesture comprehension also develops during early childhood. Research indicates that, in English- and Catalan-speaking young children, beat gesture comprehension precedes beat gesture production developmentally and suggests that beat gesture comprehension may benefit development of other cognitive skills during early childhood (e.g. memory for lists of words and information from discourse, encoding of spatial information; Austin & Sweller, 2014; Igalada et al., 2017; Llanes-Coromina et al., 2018; Macoun & Sweller, 2016; Vilà-Giménez et al., 2019). For example, 3–5-year-old Catalan-speaking children recall significantly more target words from a series in spoken
discourse when they are presented with beat gesture than when they are presented without beat gesture. However, the same research indicates that their recall of non-target words in spoken discourse is not affected by beat gesture, suggesting that beat gesture highlights words or phrases that it accompanies (i.e. lexical affiliates) rather than other unrelated words or phrases (Igualada et al., 2017). Overall, evidence from production and comprehension indicates that young children begin to produce and comprehend beat gestures at a young age, with their use of it increasing during early childhood.

To date, few studies have examined whether the development of beat gesture comprehension and production in early childhood differs by sex. This is somewhat surprising in light of evidence that girls’ early language development outpaces boys’ in several respects, including vocabulary and syntactic complexity (Bouchard et al., 2009; Fenson et al., 1994), and that this discrepancy persists into early childhood, with 3–6-year-old girls showing overall superiority and boys showing greater variability in vocabulary, grammar, speech comprehension, pronunciation, and processing of sentences and nonce words (Lange et al., 2016). Sex differences favoring girls have also been found in communicative gesture production (Eriksson et al., 2012), production of gesture-speech combinations (Özçalışkan & Goldin-Meadow, 2010), and referential gesture comprehension in the context of word learning (Foran et al., 2022). At present, it is currently unclear whether sex differences are present in beat gesture development. Of the studies that provide evidence speaking to sex differences, three failed to find any (Austin & Sweller, 2014; Colletta et al., 2010; Igualada et al., 2017), one only examined boys (Nicoladis et al., 1999), and one provided unclear evidence as sex was recorded but its effect was not formally analyzed (Mathew et al., 2018). Together, these findings highlight the importance of probing for sex differences in beat gesture development more systematically, particularly in relation to other aspects of language development.

**Beat gesture and pitch accenting relatedness**

Gestural and acoustic prominence cues are closely related in both production and comprehension, and evidence suggests that this relationship is present very early on in development. Indeed, infants exposed to Catalan as young as 9 months old can detect temporal misalignment of deictic gestures and stress (Esteve-Gibert et al., 2015). Sensitivity to the temporal alignment of gestural and acoustic prominence persists into adulthood. When Catalan-speaking adults produce deictic gestures in conjunction with stress, the apices of their gestures are temporally aligned with intonation peaks (Esteve-Gibert & Prieto, 2013). However, other components of gesture, such as extension and retraction phases, are not clearly tied to events in the speech stream such as intonation peaks. As is the case for deictic gestures and stress, in English-speaking adults, the apices of beat gestures and the intonational peaks of pitch accented words are the most consistently temporally aligned (Leonard & Cummins, 2011). Moreover, words accompanied by beat gestures are more likely to be perceived as pitch accented by German-speaking adults than words unaccompanied by beat gestures (Krahmer & Swerts, 2007).

Findings from event-related potentials (ERPs), electrophysiological brain responses to stimuli reflecting cognitive processes such as attentional and semantic processing, provide evidence demonstrating that co-occurring beat gesture and pitch accenting convey
prominence in discourse. For example, when viewing videos of a dialogue containing contrasting alternatives, the P300 ERP, which is associated with attentional processing, occurred earlier for contrastively accented alternatives in Dutch accompanied by beat gesture than for alternatives lacking contrastive accenting accompanied by beat gesture, suggesting that these cues are integrated and that attentional resources are consumed when beat gestures co-occur with non-contrastively accented alternatives (Dimitrova et al., 2016). Pitch accenting and beat gesture also affect semantic processing during language comprehension, as evidenced by the N400 component. However, Dutch presentationally accented words accompanied by beat gesture fail to elicit an interactive effect on the N400, suggesting that although these cues may capture attention when they are incorporated, they may influence semantic processing separately (Wang & Chu, 2013). The N400 component also varies as a function of the temporal alignment of beat gesture and presentational accenting in English (Morett, Landi, et al., 2020), suggesting that the temporal alignment of these cues affects semantic processing, as well. Indeed, the N400 increases in amplitude and its onset occurs earlier when beat gesture and presentational accenting are temporally asynchronous. Taken together, these results suggest that listeners are sensitive to both co-occurrence and temporal asynchrony of beat gesture and pitch accenting.

With respect to contrast, co-occurrence of beat gesture and contrastive accenting enhances processing and memory of contrastive alternatives. In Catalan discourses consisting of a passage in which two contrasting alternatives are presented (e.g. *fish shop* and *grocery shop*) followed by a passage in which one alternative is selected (e.g. *fish shop*), recognition of selected alternatives initially accompanied by beat gesture and contrastive accenting is superior to selected alternatives initially unaccompanied by both of these cues in both adults and young children (Kushch & Prieto, 2016; Llanes-Coromina et al., 2018). Furthermore, in similar English discourses, enhancement of adults’ memory for contrastive alternatives by contrastive accenting is negated when beat gesture is absent in a discursive context in which it sometimes occurs, indicating that co-occurrence of beat gesture and contrastive pitch accenting improves subsequent recognition of contrastive alternatives (Morett & Fraundorf, 2019). During comprehension of English spoken discourses referring a visual array of objects contrasting in specific features (e.g. *Click on the red star. / Now click on the blue star*), co-occurrence of contrastive accenting and beat gesture with words in similar filler discourses influenced adults’ eye movements and pupil size, indicating that co-occurrence of these cues with contrasting alternatives increases their interpretation as cues to contrast (Morett et al., 2021; Morett, Roche, et al., 2020). Together, these findings suggest that contrastive accenting and beat gesture may be most powerful as cues to contrast when they co-occur and consistently accompany contrastive alternatives.

**Emphasis cues and word learning**

Word learning is a critical aspect of language development in children. To date, many theories and models have been proposed to characterize how young children learn words and how this relates to overall language acquisition. The Emergentist Coalition Model (ECM), initially proposed by Hirsh-Pasek et al. (2000), maintains that young children use
several different inputs (e.g. perceptual, social and grammatical) to learn words in a multitude of contexts. More recently, probabilistic learning theories have been posited to explain how children use general knowledge and several other factors either directly or indirectly linked to word learning contexts to acquire language (Christiansen et al., 2005; Pace et al., 2016). These theories propose that young children integrate multiple cues to track and distinguish between different properties of words (Christiansen & Monaghan, 2006). An example of this can be seen when young children use familiarity to distinguish between novel words. When a novel object is presented among a group of familiar objects and 28–36-month-old English-speaking children are asked to find an object associated with a novel word (e.g. prilp), children typically pick the unfamiliar object (Clark, 1990; Golinkoff et al., 1992; Markman, 1994; Markman & Wachtel, 1988; Merriman et al., 1989). Furthermore, when toddlers exposed to English are presented with two objects, one of which is labeled using a common, well-known word and one of which is labeled with a novel word that represents a descriptive aspect of the object (i.e. color, shape, texture), children accurately identify the novel word and, often, can provide correct words for that novel word (e.g. fibrous vs fuzzy; Heibeck & Markman, 1987). Thus, the idea that words have contrasting meanings (i.e. the principle of contrast; Clark, 1987) scaffolds the more general understanding that different words have different meanings.

Consistent with the ECM and probabilistic learning theories, beat gesture aids word learning both individually and conjointly with pitch accenting. In English- and Catalan-speaking adults and children, observing beat gesture when learning novel words enhances subsequent memory for them (Igualada et al., 2017; Kushch et al., 2018; Morett, 2014; So et al., 2012), although some work suggests that beat gesture may facilitate memory for novel words in English-speaking adults but not children (So et al., 2012). In cases in which observing beat gesture improves Catalan-speaking children’s memory for novel words, it does not similarly improve their memory for adjacent words lacking beat gesture, indicating that beat gesture’s enhancement of memory may be localized to specific words and phrases it accompanies (Igualada et al., 2017). In Catalan-speaking adults, beat gesture’s enhancement of memory for novel words is boosted by the presence of pitch accenting (Kushch et al., 2018), suggesting that co-occurrence of these cues may similarly boost children’s word learning. In combination with ECM and probabilistic learning theories, the principle of contrast suggests that contrastive accenting and beat gesture may enhance acquisition of words both individually and conjointly. This prediction is consistent with evidence that co-occurring contrastive accenting and beat gesture enhance memory for contrasting alternatives from Catalan spoken discourse in 3–5-year-old children (Llanes-Coromina et al., 2018). To date, however, the impact of co-occurring pitch accenting and beat gesture on young children’s novel word learning in general has not yet been examined, nor have the independent and conjoint effects of contrastive accenting and beat gesture on young children’s acquisition of words, specifically.

**Current study**

The current study sought to investigate how hearing pitch accenting and observing beat gesture affect young children’s word learning, focusing in particular on the effects of hearing contrastive accenting and observing beat gesture on acquisition of contrasting
novel words with different meanings. To achieve this aim, beat gesture and contrastive accenting were independently manipulated in conjunction with contrasting novel words at learning, and young children’s knowledge of the meanings of these words was probed immediately thereafter. In addition, this study investigated developmental differences in the interpretation of beat gesture and contrastive accenting supporting the acquisition of words in young children. Although young children can successfully comprehend beat gestures between 3 and 5 years of age (Austin & Sweller, 2014; Igualada et al., 2017; Llanes-Coromina et al., 2018; Macoun & Sweller, 2016; Vilà-Giménez et al., 2019) and pitch accenting around 5 years of age (Ito et al., 2012, 2014; Lee & Snedeker, 2016; Sekerina & Trueswell, 2012; Wells et al., 2004), it is unclear whether and how their interpretation of these cues develops between the ages of 3 and 5 years. In addition to investigating the effects of beat gesture, contrastive accenting, and age, sex was included as an additional factor given that the language development of girls generally outpaces that of boys (Gleason & Ely, 2002) and that toddler girls are more likely to name labels of objects learned with iconic gestures than boys (Foran et al., 2022).

Based on the ECM, probabilistic learning theories, and the principle of contrast, and in line with previously observed beneficial effects of hearing beat gesture and observing contrastive accenting on young children’s memory for contrastive alternatives from discourse (Llanes-Coromina et al., 2018), it was predicted that novel words learned while hearing contrastive accenting or observing beat gesture would be better remembered by young children than novel words learned without emphasis cues. In addition, we hypothesized that the co-occurrence of beat gesture and contrastive accenting would further improve young children’s memory for novel words. It was predicted that these effects of contrastive accenting and beat gesture would increase with age. However, since previous research suggests that children comprehend beat gesture earlier than pitch accent (Austin & Sweller, 2014; Igualada et al., 2017; Ito et al., 2012, 2014; Lee & Snedeker, 2016; Llanes-Coromina et al., 2018; Macoun & Sweller, 2016; Sekerina & Trueswell, 2012; Vilà-Giménez et al., 2019; Wells et al., 2004), a stronger effect of beat gesture than contrastive accenting was expected in 3- and 4-year-olds. Finally, we predicted that beat gesture would facilitate contrastive word learning in girls to a greater extent than it would in boys, given evidence that girls’ use of referential gesture develops earlier than boys’ (Eriksson et al., 2012; Foran et al., 2022; Özçalışkan & Goldin-Meadow, 2010).

**Method**

**Participants**

Initially, 57 children who were 3–5 years old were recruited for the study. Three children were excluded because of technical issues that occurred during data collection, and one child completed the task twice, so the second set of data collected from them was excluded. Thus, the final sample of the study comprised 53 children who were 3–5 years old (age: $M=4.41$, $SD=0.63$; sex: 24 female and 29 male). All participants were enrolled in pre-schools or daycares in the Tuscaloosa or Gadsden areas of Alabama, USA. Parents of participants were informed about the experiment and asked to sign a consent form to
allow their child’s participation. The University of Alabama Institutional Review Board provided ethical approval for the study.

**Materials**

Experimental materials consisted of video recordings in which different actors (2 male and 2 female) picked three novel objects out of a box in succession. Each novel object was a household item that 3–5-year-old children are typically unable to name (e.g. sink stopper, T-joint). Objects were chosen based on previous studies that identified similar objects as being unknown to children aged 2–6 years old (e.g. Hall et al., 1993; Rohde & Tiefenthal, 2000).

Novel words were generated using Wuggy (Keuleers & Brysbaert, 2010) with English phonotactics and a maximum of four phonological neighbors. All novel words were one syllable, contained between 4 and 6 letters, and were consistent with English phonotactic rules. There were eight novel words comprising four pairs of two words each (e.g. clarg – prilp; see Appendix 1). There were also eight experimental objects that were sorted into four pairs of two objects each (e.g. sink stopper – T-joint; see Appendix 1). Words and object pairings were counterbalanced, such that the word clarg was sometimes assigned to the sink stopper and sometimes the T-joint. Four additional objects (one per video) were included as control objects that were never labeled to reduce guessing among labeled objects.

Four adult native English-speaking research assistants (2 male and 2 female) featured in audio stimuli were trained to produce contrastive and presentational accenting on novel words. Training consisted of the first author demonstrating each type of pitch accenting and providing feedback. Research assistants featured in audio stimuli were provided with the following script, with novel words substituted as needed: ‘Wow! I’ve always wanted a kroosk! Oh, but not a fesp’. Each novel word was produced with contrastive and non-contrastive (presentational) accenting in different recordings. In addition, research assistants featured in audio stimuli were also recorded saying the filler phrase ‘Hmm, look at that’ without varying pitch accenting in reference to the control object. All audio recordings were captured in a sound booth using professional audio recording equipment to ensure high-quality audio.

To verify that novel words with contrastive and non-contrastive pitch accenting differed acoustically, we compared their acoustic properties across words and speakers using a series of paired-sample t-tests. Table 1 presents means and standard deviations of acoustic measures along with these comparisons. Consistent with previous research (Morett et al., 2021; Morett, Roche, et al., 2020), novel words with contrastive and non-contrastive pitch accenting differed significantly on all measures except $F_0$ difference.

Four additional adult research assistants (two male and two female) not featured in audio stimuli were featured in videos. Research assistants featured in videos were trained to produce beat gestures on novel words. Training consisted of the first author demonstrating one of the most common beat gestures talkers produce in natural conversation: a rapid single downward stroke of the hand with the palm open upward (McNeill, 1992), consistent with previous research examining beat gesture as a cue to contrast (Morett et al., 2021; Morett & Fraundorf, 2019; Morett, Roche, et al., 2020), and providing
feedback. Research assistants featured in videos lip synced to sentences in recorded audio stimuli. Separate research assistants were featured in video and audio stimuli to ensure that contrastive and non-contrastive pitch accenting differed significantly. As novel objects were mentioned in sentences, research assistants featured in videos removed them from a box with their dominant (right) hand and either produced a beat gesture with their non-dominant (left) hand or kept it still. All video recordings were captured using a professional digital video recording equipment.

In post-production, intensity was normalized to the average across audio recordings within each pitch accent condition using Praat (Boersma & Weenink, 2016) to eliminate any spurious differences in intensity across individual recordings. Using Adobe Premiere, audio tracks captured during video recording were deleted, and audio recordings were temporally aligned with videos based on mouth movements and, where present, beat gestures. Videos were then trimmed to the length of audio recordings.

In a within-participants experimental design, the presence or absence of beat gesture, in combination with the presence or absence of contrastive pitch accenting, created four emphasis conditions in which the labeling of each of the experimental objects coincided with (a) beat gesture, (b) contrastive accenting, (c) both cues, or (d) neither cue (see Figure 1). Each sequence (learning trial) consisted of two objects and corresponding labels presented in two different emphasis conditions plus the filler item, which occurred either before or after the experimental items (see Figure 2). A total of four sequences were presented to each child, ensuring that both the first and second objects were presented in each learning condition and that no learning condition was repeated within a sequence. Assignment of objects and labels to each emphasis condition and speaker/model, order of presentation, and filler phrase placement were counterbalanced across participants.

### Procedure

Data were collected from all children individually in a quiet classroom at their school or daycare. Age and sex were based on data reported by the school or daycare, which originated from caregivers. Each experimental session lasted approximately 15 minutes. Caregiver consent and participant assent were obtained prior to data collection and attention was monitored throughout the experiment. Participants were told they would be playing a game in which they would be shown videos with different objects and to pay close attention.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Non-contrastive</th>
<th>Contrastive</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity (dB)</td>
<td>58.84 (5.91)</td>
<td>62.95 (0.07)</td>
<td>510</td>
<td>-8.57</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>Duration (s)</td>
<td>0.40 (0.07)</td>
<td>0.59 (0.10)</td>
<td>510</td>
<td>-25.59</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>Maximum $F_0$ (Hz)</td>
<td>319.46 (167.23)</td>
<td>362.93 (152.78)</td>
<td>510</td>
<td>-3.07</td>
<td>.002***</td>
</tr>
<tr>
<td>$F_0$ difference (Hz)</td>
<td>176.20 (169.14)</td>
<td>188.84 (149.59)</td>
<td>510</td>
<td>-0.90</td>
<td>.37</td>
</tr>
<tr>
<td>Mean $F_0$ (Hz)</td>
<td>202.28 (69.41)</td>
<td>235.67 (82.43)</td>
<td>510</td>
<td>-4.96</td>
<td>&lt;.001***</td>
</tr>
</tbody>
</table>

***p < .001; **p < .01.
After completing practice trials and indicating they were ready to begin the experimen-
tal session, participants were presented with learning videos on a laptop via a paradigm programmed in OpenSesame (Mathôt et al., 2012). After each learning video, images of the three novel objects were presented on screen, and participants were asked to identify each one by touching the correct image in reference to its corresponding word (e.g. ‘Can you touch the picture of the kroosk? Can you touch the picture of the fesp?’; see Figure 3). In between trials, children completed a brief filler task in which they were shown pictures of familiar scenes (e.g. a zoo), and asked to find an object within the picture (e.g. ‘Here is a picture of a zoo. Have you ever been to the zoo? Can you find the zebra?’) to clear their working memory. After participants completed all four trials and three intervening filler tasks, they were dismissed. Object and trial order were counterbalanced across participants. Emphasis cues were counterbalanced across trials such that all four possible cue combinations were presented to each participant twice across word pairs.

Figure 1. Emphasis Conditions.
Results

Response accuracy data from the referent identification task were analyzed using logit mixed effect models, which model the log odds (or logit) of a correct response on each trial (1 = correct, 0 = incorrect) without assumptions of linearity, normality, or homoscedasticity. In these models, specified factors were included as fixed effects, and participants and novel words were included as random effects, with referent identification accuracy as the outcome variable. All continuous fixed effects were mean centered, and all categorical fixed effects were coded using mean centered (Helmert) contrast coding. All models included random intercepts only as models with random slopes of factors by participants either failed to account for significant variance above them or failed to converge. All models were fit in R using the glmer() function of the lme4 package (Bates et al., 2015) and null hypothesis significance testing was conducted using the lmerTest package (Kuznetsova et al., 2017). All data and analysis code are publicly available on the Open Science Framework (Morett et al., 2022).

Our main analysis examined whether referent identification accuracy differed based on the presence of beat gesture and contrastive accenting at learning as well as age and
Table 2. Fixed effect (top) and variance estimates (bottom) in log odds for multi-level logit model of referent identification accuracy by beat gesture, contrastive accenting, age, and sex (observations = 424).

<table>
<thead>
<tr>
<th>Fixed effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>Wald z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−1.02</td>
<td>0.83</td>
<td>−1.22</td>
<td>.22</td>
</tr>
<tr>
<td>Beat gesture</td>
<td>1.52</td>
<td>0.78</td>
<td>1.95</td>
<td>.05†</td>
</tr>
<tr>
<td>Contrastive accenting</td>
<td>0.23</td>
<td>0.78</td>
<td>0.30</td>
<td>.77</td>
</tr>
<tr>
<td>Age (years)</td>
<td>0.27</td>
<td>0.19</td>
<td>1.45</td>
<td>.15</td>
</tr>
<tr>
<td>Sex</td>
<td>−1.88</td>
<td>0.83</td>
<td>−2.25</td>
<td>.02‡</td>
</tr>
<tr>
<td>Beat gesture × contrastive accenting</td>
<td>0.78</td>
<td>0.78</td>
<td>1.00</td>
<td>.32</td>
</tr>
<tr>
<td>Beat gesture × age (years)</td>
<td>−0.31</td>
<td>0.17</td>
<td>−1.79</td>
<td>.07‡</td>
</tr>
<tr>
<td>Contrastive accenting × age (years)</td>
<td>−0.04</td>
<td>0.17</td>
<td>−0.23</td>
<td>.82</td>
</tr>
<tr>
<td>Beat gesture × sex</td>
<td>2.79</td>
<td>0.79</td>
<td>3.55</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>Contrastive accenting × sex</td>
<td>−0.88</td>
<td>0.79</td>
<td>−1.13</td>
<td>.26</td>
</tr>
<tr>
<td>Age × sex</td>
<td>0.38</td>
<td>0.19</td>
<td>2.04</td>
<td>.04*</td>
</tr>
<tr>
<td>Beat gesture × contrastive accenting × age (years)</td>
<td>−0.18</td>
<td>0.17</td>
<td>−1.03</td>
<td>.31</td>
</tr>
<tr>
<td>Beat gesture × contrastive accenting × sex</td>
<td>0.37</td>
<td>0.78</td>
<td>0.47</td>
<td>.64</td>
</tr>
<tr>
<td>Beat gesture × age (years) × sex</td>
<td>−0.58</td>
<td>0.18</td>
<td>−3.33</td>
<td>&lt;.001***</td>
</tr>
<tr>
<td>Contrastive accenting × age (years) × sex</td>
<td>0.19</td>
<td>0.18</td>
<td>1.08</td>
<td>.28</td>
</tr>
<tr>
<td>Beat gesture × contrastive accenting × age (years)× sex</td>
<td>−0.04</td>
<td>0.17</td>
<td>−0.25</td>
<td>.81</td>
</tr>
</tbody>
</table>

Random effect  

<table>
<thead>
<tr>
<th>s²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant</td>
</tr>
<tr>
<td>Word</td>
</tr>
</tbody>
</table>

***p < .001; *p < .05; †p < .10.

sex. Table 2 displays parameter estimates for the model. We observed a marginal main effect of beat gesture, indicating that the referents of words learned with beat gesture were identified slightly more accurately (M=0.57, SD=0.50) than the referents of words learned without beat gesture (M=0.52, SD=0.50). There was no main effect of contrastive accenting, indicating that the referents of words learned with contrastive accenting (M=0.57, SD=0.50) were identified no more accurately than the referents of words learned without contrastive accenting (M=0.53, SD=0.50). We observed a significant main effect of sex, indicating that the referents of contrasting novel words were identified with marginally greater accuracy by girls (M=0.59, SD=0.49) than by boys (M=0.51, SD=0.50). We also observed significant interactions of gesture by sex, age by sex, and gesture by sex by age, as well as a marginal interaction of gesture by age. None of the interactions including contrastive accenting reached significance, however. Simple effect analyses by gesture and sex indicated that referent identification accuracy increased significantly with age only for novel words learned without beat gestures for boys (B=0.32, SE=0.08, t=4.11, p < .001), whereas it did not differ by age for novel words learned with beat gestures for boys (B=−0.06, SE=0.08, t=−0.75, p=.46) or regardless of whether novel words were learned with beat gesture (B=0.04, SE=0.09, t=0.48, p=.63) or without it (B=−0.09, SE=0.08, t=−1.03, p=.31) for girls (see Figure 4).
Given that boys’ identification of the referents of novel words learned without beat gesture increases from below chance and below their identification of the referents of novel words learned with beat gesture at age 3 years to above chance and comparable to their identification of words learned with beat gesture at age 5 years, this finding suggests that beat gesture may provide an additional cue to contrast that helps boys learn the meanings of contrasting novel words with success similar to girls at 3 years of age.

We also conducted supplementary analyses exploring the influences of trial number, novel word order at learning, referent order at testing, and unnamed object order on referent identification accuracy. We observed a significant negative main effect of trial number, indicating that referent identification accuracy decreased across trials ($B=-0.19$, $SE=0.09$, $z=-2.06$, $p=.04$). These analyses revealed no significant main effects of learning order ($B=-0.20$, $SE=0.20$, $z=-1.00$, $p=.32$) or testing order ($B=-0.08$, $SE=0.20$, $z=-0.39$, $p=.70$) but revealed a significant main effect of unnamed object order ($B=-0.22$, $SE=0.10$, $z=-2.16$, $p=.03$), indicating that referent identification accuracy was higher when unnamed objects were presented before novel words ($M=0.60$, $SD=0.49$) than after them ($M=0.50$, $SD=0.50$).

**Discussion**

This research investigated how observing beat gesture and hearing contrastive pitch accenting affect young children’s learning of novel words contrasting in meaning. Inconsistent with our hypotheses, the results revealed only a non-significant trend of beat gesture and a marginal interaction between it and age; no main effects or interactions of contrastive accenting were found. However, the results revealed that 3-year-old boys learn novel words with contrasting meanings accompanied by beat gesture more effectively than novel words unaccompanied by beat gesture, but this effect decreased in 4–5-year-old boys as they became more adept at learning these words. Word learning in girls, by comparison, did not benefit from observing beat gesture. These findings suggest
that young children’s ability to leverage beat gesture as a cue to contrast in the context of word learning may differ by sex, particularly in combination with age.

Aside from a marginal main effect of beat gesture, the results failed to reveal any overall main effects or interactions of hearing contrastive accenting and observing beat gesture. With respect to contrastive accenting, previous research has shown that additional support from discourse is needed for young children to recognize discourse with contrastive accenting (Ito et al., 2014; Sekerina & Trueswell, 2012; Wells et al., 2004), and the only other cue to contrast available in discourse in the current study was the words ‘but not’. Moreover, a key difference between the current study and previous research showing effects of these cues in early childhood (Austin & Sweller, 2014; Igualada et al., 2017; Lee & Snedeker, 2016; Llanes-Coromina et al., 2018; Macoun & Sweller, 2016; Vilà-Giménez et al., 2019) and adulthood (Kushch et al., 2018; Morett et al., 2021; Morett & Fraundorf, 2019; Morett, Roche, et al., 2020) is that the current study examined their effects in the context of a novel word learning task. Unlike the discourse memory tasks used in many previous studies, which require information from a coherent discourse to be remembered, the novel word learning task requires the phonological forms of novel words to be mapped onto their referents and for this mapping to be remembered. Although pairs of words were presented in contrast to one another similar to information in many of these discourses, the difference between word learning and discourse comprehension, which may have been greater for young children, may explain why beat gesture and contrastive accenting were less effective as cues to contrast in the current study of word learning than they were in previous research on discourse memory.

Moreover, the results failed to reveal an overall main effect of age, and aside from a marginal interaction of age with beat gesture, any interactions of it with beat gesture or contrastive accenting alone. Regarding the lack of an overall main effect of age, this may be due to the competency of the majority of 3–5-year-olds we tested on the novel referent identification task, as evidenced by their above-chance accuracy. Indeed, our previous work using a similar novel word learning paradigm with a comparably aged sample failed to show an overall effect of age, with the majority of children showing above-chance accuracy, although interactions between age and the factors of interest were found (Stager et al., 2022). The absence of interactions between age and contrastive accenting and beat gesture in the current study was more surprising in light of research demonstrating that production of beat gesture increases during early childhood (Colletta et al., 2010; Mathew et al., 2018) and that interpretation of it does not mature until adulthood (Austin & Sweller, 2014; Ito et al., 2012, 2014; So et al., 2012), as well as work showing that young children’s interpretation of contrastive accenting is immature (Ito et al., 2014; Sekerina & Trueswell, 2012; Wells et al., 2004). However, it is worth noting that many existing studies examining young children’s interpretation of these cues have not reported age effects (Austin & Sweller, 2014; Igualada et al., 2017; Ito et al., 2012, 2014; Lee & Snedeker, 2016; Llanes-Coromina et al., 2018; Macoun & Sweller, 2016; Vilà-Giménez et al., 2019). Thus, the lack of overall age effects and interactions of age with beat gesture and contrastive pitch accenting suggests that the methods used in the current study to ascertain young children’s ability to leverage beat gesture and contrastive accenting to facilitate the learning of novel words with contrastive meanings may be insufficiently sensitive to any developmental changes occurring during early childhood.
The findings of the current study add to the literature on the comprehension of emphasis cues in early childhood by demonstrating that interpretation of beat gesture in a contrastive word learning task varies with age by sex. Previous work suggests that girls’ language development in early childhood outpaces that of boys in several respects (Lange et al., 2016), and that infant and toddler girls produce more types of referential gestures and gesture-word combinations than boys (Eriksson et al., 2012; Germain et al., 2022; Özçalışkan & Goldin-Meadow, 2010) and are more likely to name objects learned with referential gestures than boys (Foran et al., 2022). With respect to non-referential beat gesture development, the results of previous studies fail to provide evidence of sex differences (Austin & Sweller, 2014; Colletta et al., 2010; Igualada et al., 2017; Mathew et al., 2018). By comparison, the results of the current study show that, at age 3, boys leverage beat gesture to enhance their learning of novel words with contrasting meanings to the point that it is comparable to that of girls, whereas by age 4–5 years, boys no longer do so. These findings parallel findings for children with developmental language disorder, which disproportionately affects boys, showing that they rely on declarative memory to compensate for differences in procedural memory (Ullman & Pierpont, 2005), allowing them to learn novel inflected verbs effectively (Harmon et al., 2022). Thus, these sex differences in the impact of beat gesture may have been influenced by sex differences in word learning favoring girls in early childhood (Lange et al., 2016). This may be the case given that previous studies of beat gesture not showing sex differences examined either beat gesture production (Colletta et al., 2010; Mathew et al., 2018) or the impact of observing beat gesture on the memory for known words or discourse. Thus, beat gesture may serve as a compensatory cue to contrast for 3-year-old boys, providing them with information about contrasting meanings that they have difficulty gleaning from speech. If this explanation is accurate, the sex differences observed in the current study may be restricted to the use of beat gesture to convey contrast in the context of novel contrastive word learning. This can be explored in future research by examining the effects of observing beat gesture on 3–5-year-old boys’ and girls’ memory for known words with contrasting meanings with their learning of novel words with non-contrasting meanings.

One limitation of the current study is the sample size and composition. Because data were collected in person, commencing before the coronavirus disease-19 pandemic when Internet-based data collection from children became mainstream, and because we wanted to avoid switching data collection modalities mid-study, our means for recruitment were limited. Future research would do well to take advantage of the resources for and interest in Internet-based data collection to determine the extent to which the findings are replicable in this modality with larger and more diverse samples. Relatedly, another limitation of the current study is the number of trials and items. Although they are not unusual for studies of language development in young children, including additional items, if possible, would increase the reliability of the findings. It is worth noting that each child completed only one learning trial and two test trials immediately thereafter for each pair of novel words. Although fast mapping research has shown that children are capable of single trial word learning (Heibeck & Markman, 1987), evidence suggests that repetition of novel words, particularly within the same context, may lead to more effective learning (Horst, 2013). An additional limitation of the study was data collection
using only a referent identification task. Although pilot testing indicated that children struggled to produce novel words on demand, consistent with Foran et al. (2022), inclusion of a novel word production task in addition to a referent identification task would yield additional data providing insights into the effects of beat gesture and contrastive accenting on novel word learning in early childhood. Finally, the word learning task used in the current study had limited discursive context, and novel word knowledge was only tested immediately following learning. Thus, it is possible that the effects of beat gesture and contrastive accenting on novel contrastive word learning may be evident in more extensive discursive contexts (Llanes-Coromina et al., 2018) and following longer delays (De Nooijer et al., 2013).

In conclusion, our findings revealed that although neither observing beat gesture nor hearing contrastive accenting enhances the learning of novel contrastive words for all young children, 3-year-old boys are better able to learn novel words with contrasting meanings accompanied by beat gestures than unaccompanied by beat gestures. Thus, they may leverage beat gesture as a cue to contrast, compensating for difficulties deducing contrast from speech alone, and they may rely on it less as their ability to deduce contrast from speech increases during early childhood. This finding suggests that beat gesture may serve as a visual cue to contrast scaffolding young children’s learning of words with contrasting meanings by supplementing cues to contrast conveyed via speech.

Acknowledgements
This work was presented at the 2021 meeting of the Psychonomic Society and the 2022 meeting of the International Society for Gesture Studies. The authors thank Mona Anchan, Brian Rivera, and Nathaniel Shannon IV for assistance with video recording of stimuli; Ellen Macaruso and Wesley Webber for assistance with audio recording of stimuli; and Carlee Brandon, Raymond Bulger, Margaret Doheny, Gracie Springer, and Jacob Zuluaga for assistance with paradigm preparation and/or data collection.

Author contributions
Laura M. Morett: Conceptualization; Formal analysis; Methodology; Project administration; Supervision; Visualization; Writing – original draft; Writing – review & editing.
Cailee M. Nelson: Data curation; Investigation; Methodology; Writing – original draft; Writing – review & editing.
Sarah S. Hughes-Berheim: Methodology; Resources; Software; Writing – review & editing.
Jason Scofield: Conceptualization; Methodology; Project administration; Resources; Supervision; Writing – review & editing.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD
Laura M. Morett https://orcid.org/0000-0002-1251-7213
References


Appendix 1. Novel word pairs and their corresponding novel object pairs.

<table>
<thead>
<tr>
<th>Pseudowords</th>
<th>Object description</th>
</tr>
</thead>
<tbody>
<tr>
<td>clarg/prilp</td>
<td>red canteen/gray sponge</td>
</tr>
<tr>
<td>fesp/kroosk</td>
<td>red funnel/white stopper</td>
</tr>
<tr>
<td>yozz/smolt</td>
<td>green pipe/blue colander</td>
</tr>
<tr>
<td>skeemp/rabe</td>
<td>silver filter/red T joint</td>
</tr>
</tbody>
</table>