Language Sciences 87 (2021) 101414

Contents lists available at ScienceDirect

Language Sciences

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

Show gestures direct attention to word–object relations in typically developing and Autistic Spectrum Disorder children

Nancy de Villiers Rader[a,*], Patricia Zukow-Goldring[b], Theodore Alhanti[a]

[a] *Psychology Department, Ithaca College, USA*
[b] *Center for the Study of Women, University of California, Los Angeles, USA*

ARTICLE INFO

Article history:
Available online 2 August 2021

Keywords:
Word learning
Gestures
ASD
Development

ABSTRACT

Previous research by Rader and Zukow-Goldring (2010, 2012, 2015) on word learning found that infants benefit from a speaker's use of a *show* gesture. Here we examined whether this benefit exists in typically developing (TD) children aged 24.2–76.7 months, as well as in children with Autism Spectrum Disorder (ASD) aged 5.8–8.2 years. While we recorded eye gaze, children viewed a video showing a speaker introducing two novel objects with nonce words. In the *show* condition, the speaker moved the object in synchrony with her speech, while in the static condition she held the object stationary. After the speaker introduced the objects, we assessed word learning. Analysis of word learning found that the TD toddlers and the ASD children benefited from the *show* condition, while preschool and elementary-aged TD children learned equally well in both gesture conditions. We propose that *show* gestures present metamodal information specifying word reference relationships and that this information is used by TD toddlers and older ASD children in learning words.

A primary puzzle in early language learning is how the relationship between word and referent is learned, especially before infants are skilled at detecting another's target of attention through pointing or eye gaze. Based on her observations of English and Spanish speakers from the western U.S. and rural and urban Spanish speakers in Central Mexico, Zukow-Goldring (1997) proposed that caregivers embed their communications in multi-sensory regularities. She noticed that speakers often loom and retract an object and, as a word is said, rotate the object around the axis of the hand/arm from *en face* to the side and back. In this way, the looming-retraction and the movement to the side and back match the onset/offset, rhythm, tempo, and duration of the saying of the word. Zukow-Goldring, in describing this type of gesture, called it a *show* gesture. Zukow-Goldring and Rader (2001) hypothesized that this synchrony of gesture and speech assists early word learning, making use of a rhythmic co-occurrence across seeing and hearing. In this paper, we will refer to this co-occurrence as creating metamodal information.

The term "information" has had many interpretations even within ecological psychology theory. A very general definition was used by Reed (1996) who wrote in defining "information", "Information is something that informs about something else." (p48). Gibson (1966) restricts the meaning somewhat by writing that information about something means specificity to something. This meaning is incorporated in a central tenet of Gibsonian theory positing that the structure over time in a sensory array specifies properties of the environment, allowing for the perception of affordances by an observer, where an affordance is an aspect of the organism–environment relationship providing a basis for adaptive behavior by an organism in

* Corresponding author. Department of Psychology, Ithaca College, Ithaca, NY, 14850, USA.
E-mail address: rader@ithaca.edu (N. de Villiers Rader).

https://doi.org/10.1016/j.langsci.2021.101414

an environment. This meaning of "information" does not lead to the notion of perceiving information, but, rather, picking up or resonating to information for the purpose of perceiving meaningful aspects of the environment by an observer.

From an ecological psychology perspective, synchrony of gesture and speech can be seen as functioning as a type of metamodal information, where the co-occurrence of visual and auditory information across time specifies the word-object pairing. The information is in neither the visual nor the auditory sensory arrays, but in the temporal co-occurrence across the two modalities. The pick-up of information is, thus, at a "meta" level and is not merely multimodal (cf. Read and Szokolszky, 2018). That this metamodal information can be present given a speaker's creation of it is undeniable, and using 3-D motion analysis Sahlstrom and Rader (2009) have documented its spontaneous use by mothers while introducing a novel object to their 12-month-old infants. However, whether this metamodal information is picked up and resonated to for the purpose of perceiving the word-object pairing is a critical question for understanding early word learning.

Research with infants 9–15 months of age (Rader and Zukow-Goldring, 2010, 2012, 2015) found evidence that *show* gestures effectively contribute to infant word learning and pointed towards a means by which this happens. Using eye tracking technology, they found that *show* gestures brought the infants' attention to the object as the referring word was uttered and that pupil dilation occurred at that time. Laeng et al. (2012) link pupil dilation with better attentional processing associated with arousal. Rader and Zukow-Goldring (2015) conclude that the speaker's orchestration of infant attention by means of *show* gestures not only directs an infant's gaze to an object as the word which refers to it is uttered, but also pumps up arousal just at the critical time. They inferred that the synchrony present in a *show* gesture across gesture and speech takes advantage of the natural attentional tendencies of the infant as the result of the metamodal information provided by temporal characteristics existing across the visual and auditory modalities.

This infant research suggests that *show* gestures propel word learning in infancy through providing information existing across modalities that infants resonate to and use to perceive word-object pairings. But to what extent are *show* gestures important for older children for whom their language knowledge might decrease their need to pick up the information provided by a *show* gesture? In addition, we wondered whether *show* gestures might be important in aiding non-verbal children with Autistic Spectrum Disorder (ASD) in word learning. It is known that children with ASD have difficulty with attention (Keehn et al., 2013) and are less likely to orient towards social sounds (Dawson et al., 1998). We hypothesized that metamodal information provided by synchronous gestures during speech would assist ASD children with word learning if they indeed pick up this information.

Two experiments are described here to address these questions. In Experiment1, we examined the effects of either a *show* gesture or a static gesture on word learning in typically developing (TD) toddlers, preschoolers, and early elementary school children. In Experiment 2, we looked at how gesture type affects word learning in a small group of non-verbal ASD children. We predicted that the *show* gesture would be more likely to assist word learning in the toddlers and ASD children.

1. Method

1.1. Experiment 1

1.1.1. Participants

Participants were recruited from local area schools and day-care programs. All participants came from homes in upstate NY in which English was the primary language spoken. Our sample was identified through a parent questionnaire as about 80% Caucasian and 20% other, reflecting the local demography. There were 36 typically developing (TD) children aged 24.2 months–76.7 months; of these 20 were boys and 16 were girls. These children were divided into three groups based on their enrollment in day-care programs or in an elementary classroom – toddler, preschool, and early elementary. Ages of the participants in each of these groups were as follows: 10 toddlers aged 24.2–28.6 months ($M = 26.2$, $SD = 1.7$), 15 preschoolers aged 32.1–48.3 months ($M = 41.9$, $SD = 5.8$), and 11 early elementary children aged 49.5–76.7 months ($M = 57.5$ $SD = 9.8$). Parents of the participants filled out the short-form version of the MacArthur Communicative Development Inventories (Fenson et al., 2000), which was used to determine the participants' level of language development. The mean score for the toddlers was 68.4 ($SD = 33.6$), for the preschoolers 88.43 ($SD = 26.7$), and for the elementary children 99.82 ($SD = .41$).

1.1.2. Research design

In this mixed design study, there was one within-subject factor, Gesture Type, and one between-subjects factor, Group. All participants saw a *show* gesture and static gesture presentation; in each gesture condition one of two novel objects was presented with a corresponding nonce word. The ordering of the gestures and the nonce word used for each object was counterbalanced across participants. Following presentation of the objects and words, participants were tested on word learning.

Approval for this research, both Experiments 1 and 2, was received from the Ithaca College Institutional Review Board.

1.1.3. Materials and procedure

Informed consent was obtained from a parent or guardian for each participant prior to data collection.

A Mangold Eye Tracking System was used to collect eye-gaze data, with participants seated approximately two feet in front of a computer monitor. The light source and camera used to track eye gaze was located in a hardware strip just below the bottom of the monitor; the laptop used to collect the data was off to one side and of sight. Younger children sat in a child-size truck (see

Fig. 1) while older children sat on a chair at a table. Those children who sat in the truck were told that to be a "good driver" they had to pay good attention to the screen; this proved a successful strategy for maintaining general attention to the task.

Each subject's eye gaze position was individually calibrated. Once calibrated, the monitor displayed a video of a speaker holding a novel object and introducing it using a nonce word while using a *show* or a static gesture.

Fig. 1. Child seated in the "truck" for viewing the computer monitor.

1.1.3.1. Creation of the video word presentation segments. The videos used to present the objects and their corresponding nonce words were the same as those used by Rader and Zukow-Goldring (2010, 2012, 2015). The video segments were professionally filmed against a blue screen and then digitally manipulated so that each segment was identical except for counterbalancing and the gestures used. The base video was of a female speaker who moved the toy in rhythm with her speech, using a typical *show* gesture. The background was a simple green ground with a blue sky that had a sun and two clouds (see Fig. 2). To control for variations in speech characteristics (prosody, pitch, intensity), the video technician edited the video files so that the same "talking head" appeared in each gesture condition. To create the static gesture presentation, a portion of the video image below the neck of the speaker was clipped as a still image and pasted into the video segment used throughout that condition.

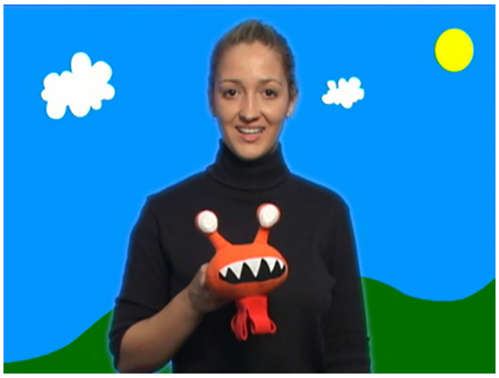

Fig. 2. Screenshot of word–object presentation segment.

1.1.3.2. Content of the word-object presentation segments. The nonce words used for the objects were *gepi* and *tano*. In half the counterbalancing tapes, *gepi* was presented with a *show* gesture, whereas for the other half *tano* was presented with a *show* gesture. Each word-object presentation segment lasted 35 s and was then repeated. For all conditions, the following script was used, where *X* is either *tano* or *gepi*: *Look at the X. See it. X. X. See it. X. Look at the X. X.* Thus, each word was presented a total of 12 times within each gesture condition.

In the *show* gesture presentation, the speaker brought the object to center as she said *Look at* or *See it*. Next, the speaker rotated the object around the vertical axis of her arm/hand to the left as she said the first syllable and then back to center in front of her upper body as she said the second. Then, when repeating the word, she loomed the object as the word was said, with the start of the looming action co-occurring with the first syllable of the word and the retraction of the hand/arm with the second syllable. In the static gesture presentation, the object was held by the speaker at a point midway between the body and the end of a looming action and it did not move at all.

1.1.4. Segments for testing word learning

Following the introduction of the two objects and their words, participants were presented with a word learning task. The segments used to test for word learning were 20 s long and were repeated once for each word. For all segments, the speaker was shown in the middle of the screen with the *gepi* to one side of her and the *tano* to the other side; no gestures were used during the word learning assessment (see Fig. 3). For half the tapes, the first request to look was for the word presented with the dynamic synchronous gesture, and for the other half, the first request was for the word presented with the static gesture. The following script was used, where *X* is either *tano* or *gepi*: *Look at the X. Are you looking at the X? Find the X*. Thus, for each word there were six requests to look at the referent object.

Fig. 3. Screenshot of word learning assessment segment.

1.2. Results

The measure of word learning used was the number of correct looks over the total number of looks to the two objects. To analyze word learning, a mixed ANCOVA was carried out with MacArthur scores as a covariate, the three groups as a between-subjects factor and gesture as a within-subjects factor. There was a significant Gesture-by-Group interaction, $F(2, 31) = 3.84$, $p = .032$, $\eta_p 2 = .198$. Follow-up one-way analyses indicated no group differences for the *show* gesture, $F(2,33) = .435$, $p = .651$, but a significant difference for the static gesture, $F\ (2,33) = 3.166$, $p = .055$. A post-hoc LSD test found significant differences for the static condition between the toddlers and the preschoolers ($p = .051$) and between the toddlers and the elementary children ($p = .024$). The means and standard deviations for each group are presented in Fig. 4 and were as follows. For the static condition, the toddler mean was .53 ($SD = .16$), the preschool mean was .72 ($SD = .27$) and the elementary-children mean was .78 ($SD = .26$). For the *show* condition, the respective means and standard deviations were $M = .67$ ($SD = .29$), $M = .66$ ($SD = .33$), and $M = .76$ ($SD = .19$).

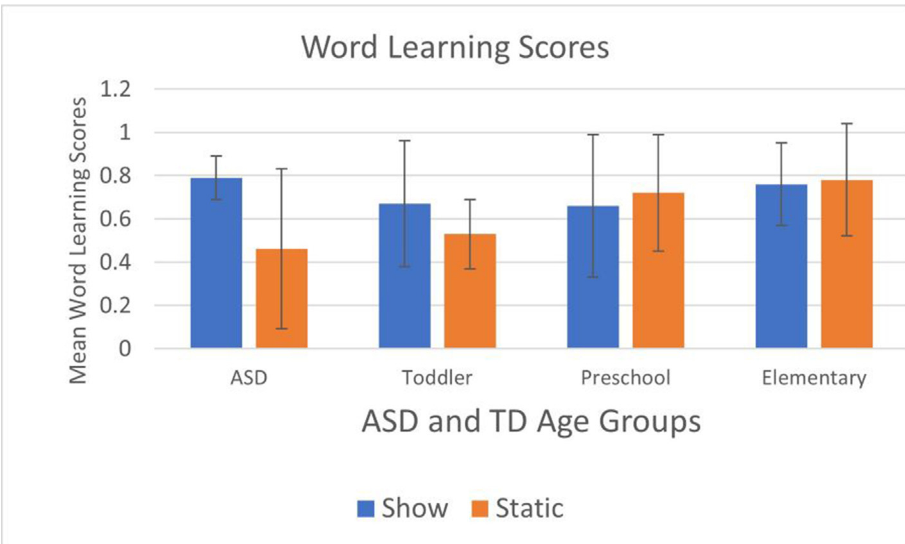

Fig. 4. Mean word learning scores for the *show* and static gesture conditions with standard deviation bars.

While the three groups were not found to differ for the *show* condition, they did differ when a static gesture was used; in that case, as shown by the post-hoc comparisons, the toddlers performed more poorly than the preschoolers or elementary children. In other words, the toddlers were able to do as well as the older children when the *show* gesture was used but not when the static gesture was used.

In addition, we found that for the youngest TD children, word learning was significantly different from chance level (.5) only for the *show* condition $t\,(9) = 1.865$, $p < .05$. For the two oldest TD age groups, word learning differed significantly from chance for both gesture conditions; for the *show* condition, $t\,(26) = 3.641$, $p = .001$ and for the static condition, $t\,(26) = 4.768$, $p < .001$.

In examining the total duration of fixations to the object as its word was uttered (FDW), an ANOVA found an overall effect of gesture type, showing less fixation time on the object during the utterance of its referent when a static gesture was used, $F\,(1,33) = 12.37$, $p = .001$, $\eta_p 2 = .273$. In msec, the *show* condition means (*sd*) were 4539 (2841) for the toddlers, 5704 (2878) for the preschoolers, and 4298 (3148) for the elementary school children. The static condition means (*sd*) were 3182 (2202) for the toddlers, 3176 (1801) for the preschoolers, and 2819 (1716) for the elementary school children (see Fig. 5). Thus, for all three TD groups of children, the *show* gesture served to orchestrate looks to the object at the critical time its referent was uttered; however, this orchestrating of attention was related to improved word learning only for the toddler group.

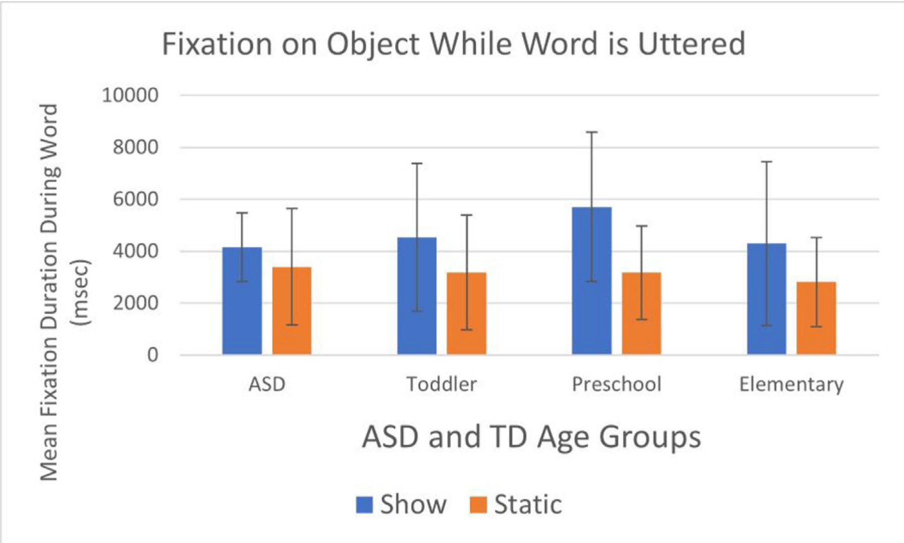

Fig. 5. Mean fixation durations during the object's word utterance for the *show* and static gesture conditions with standard deviation bars.

1.3. Experiment 2

1.3.1. Participants

For Experiment 2, there were six ASD children aged 69.5–98.4 months ($M = 83.0$, $SD = 11.2$); of these there were five boys and one girl. These participants were recruited from a special education classroom in the local school district. All participants in the ASD group had received a DSM diagnosis and were described and observed as non-verbal, exhibiting no comprehensible speech. They were also assessed by a teacher or parent using the Autism Spectrum Rating Scales (Goldstein and Naglieri, 2010); percentile scores indicated that their severity of ASD ranged between 96 and 99. Thus, while this sample was small, it was quite homogenous in terms of the severity of ASD symptoms. Two additional ASD students in the classroom could not be included as participants because we were unable to obtain a calibration for eye tracking from them.

1.3.2. Research design

As in Experiment 1.

1.3.3. Materials and procedure

Informed consent was obtained from a parent or guardian for each participant prior to data collection.

Testing was carried out at the participants' school in an unused room. The child's aide accompanied the child to the testing room and sat next to the child during testing. As in Experiment 1, a Mangold Eye-Tracking System was used to collect eye-gaze data. Participants sat in a child-sized chair approximately two feet in front of a computer monitor.

Each subject's eye-gaze position was individually calibrated. Once calibrated, the monitor displayed the video segments described in Experiment 1 for presenting the novel objects with their corresponding nonce words and for assessing word learning.

1.4. Results

For this analysis we compared the word learning results of the ASD children with the results of the TD children from Experiment 1. Because the two older groups of children in Experiment 1 did not differ, these groups were combined into one "school age" group. To analyze word learning, a Mixed ANOVA was carried out for Group (toddler, school age, and ASD), the between-subjects factor, and Gesture (*show* vs. static), the within-group factor. The ratio of correct looks to total looks during the word learning segments was again used as the dependent variable. There was a significant main effect of Gesture, $F(1, 39) = 5.48$, $p = .024$, $\eta_p 2 = .123$ and a significant interaction of Gesture by Group, $F(2, 39) = 3.69$, $p = .034$, $\eta_p 2 = .159$. Means and standard deviations are shown in Fig. 4 and were as follows: Toddler *show* $M = .669$, $SD = .29$ and Toddler static $M = .523$, $SD = .163$; School-age *show* $M = .698$, $SD = .278$, School age static $M = .746$, $SD = .263$; ASD *show* $M = .793$, $SD = .104$; ASD static $M = .462$, $SD = .368$.

To examine the interaction further, a One-Way Between-Subjects ANOVA was carried out for the *show* and for the static gesture conditions. For the *show* condition, there was no significant difference across groups, $F(2,39) = .438$, $p = .649$. For the static condition, the group difference was significant, $F(2, 39) = 4.51$, $p = .017$, $\eta_p 2 = .188$. Post-hoc Multiple Comparisons using the LSD test found significant differences between the toddler and school age children ($p = .027$) and between the ASD and school age children ($p = .021$). These results lead to the conclusion that the ASD children did not differ in a detectable way from the TD children on the word learning task when a *show* gesture was used – in fact, their score was the highest across the samples. However, the ASD children performed worse than the school-age children when a static gesture was used.

Examining the word learning scores, we find that the ASD children scored above chance level (.5) for the *show* condition, $t(5) = 6.88$, $p = .001$ but not for the static condition, $t(5) = -2.54$, $p = .81$.

In examining fixations on the object while its referent was uttered (FDW) for the ASD children compared with the TD toddlers and school age children, we carried out a Group (toddler, school age, ASD) by Gesture (*show* vs static) ANOVA with FDW as the dependent variable. There was a significant effect of gesture, $F(1,39) = 6.634$, $p = .014$, $\eta_p 2 = .145$, with greater FDWs in the *show* conditions. The interaction of Gesture and Group was not significant, $F(1,39) = .596$, $p = .556$. The ASD mean *(sd)* for the *show* condition was 4153 (1330) and 3400 (2248) for the static condition (see Fig. 5).

To determine whether the pattern of differences between the ASD children and the TD children could be due to a difference in overall attention, we analyzed the total time viewing the video screen during the word presentations using a Between-Groups ANOVA. We did not find any significant difference across Groups, $F(2, 39) = 1.307$, $p = .282$. Thus, any group differences found are not the result of differences in the total time spent looking at the video segments.

2. Discussion

Ecological psychology (e.g., J.J. Gibson, 1979, E.J. Gibson and Pick, 2000) posits that organisms perceive meaningful aspects of their environments by resonating to or discovering the invariant relations between structured input to the sensory systems and features of the environment. Pick-up of this information allows organisms to respond adaptively. When an organism picks up and resonates to the information present, it is able to perceive more of what the environment affords. In considering a developing organism, one question that arises is how resonating to information changes during development given an ever-changing organism and how it might be different for individuals with a non-typical developmental pathway.

J.J. Gibson wrote in a 1972 "Purple Peril" (mimeographed handout to his seminar) entitled, "The Affordances of the Environment", "… the affordance of something is assumed not to change as the need of the observer changes. The edibility of a substance for an animal does not depend on the hunger of the animal. The walk-on-ability of a surface exists whether of not the animal walks on it (although it is linked to the locomotor capacities of that species of animal, its action system)." For development, the question is what affordances the organism has a need for not just in a particular segment of time but as its developmental needs change. Note that it is not the case that the affordance changes but that the organism changes.

As Rader and Zukow-Goldring (2010, 2012, 2015) have shown, infants are able to pick up the metamodal information presented by a *show* gesture to support perception of the pairing of a spoken form – a nonce word – and an object. In the research reported here, we examined how development might affect the utility of that information and how children with a non-typical pattern of development might differ in its use. We found that, indeed, the presence of the metamodal information provided by a *show* gesture has differing import with development and that the ASD children in this sample differed from their typically developing counterparts in their resonance to and use of this information.

The fact that overall attention to the video segments was the same for the static and *show* gesture conditions indicates that for children who benefitted from the *show* gesture, the results are not due simply to looking longer at the video. Also, the *show* gesture led all of the TD and ASD children to fixate on the object as the utterance referring to it was produced, indicating that the gesture was guiding their eye gaze, although this directing of attention did not impact word learning in the preschool or elementary-school children. – they learned as well in the static conditions when the speech-gesture synchrony was absent. In contrast, a *show* gesture contributed to word learning in typically developing toddlers and non-verbal ASD children, i.e., they resonated to the metamodal information in perceiving the word-object pairing.

In fact, when word-object pairings were introduced with a *show* gesture, word learning by the ASD children did not differ from that of TD children aged 49–76 months (the highest level achieved by any of the TD age groups), with a score of .79 for the ASD children and .76 for the oldest TD group. By contrast, when word-object pairings were introduced with a static gesture, the ASD word learning was most similar to the youngest TD group, aged 24–28 months, with a score of .46 for the ASD children and .52 for the youngest TD group – both scores not significantly different from chance level performance.

The findings of this research should encourage ecological psychologists to consider how individual characteristics may impact the pick-up of structure in stimulation from the world and its utility for an organism. That the structure is there to be picked up is only half of the story; the other half is the nature of the organism throughout development, whether typical or atypical.

Finally, it should be noted that this research and discussion has focused on the organism side of the organism–environment relationship. As a child becomes more proficient in following attentional cues and in their knowledge of language structure, the speaker is unlikely to produce the metamodal information available in a *show* gesture. Thus, not only does the organism change with development, but the environmental information present changes as well when a speaker adapts their presentation to a child's perceived needs. Perhaps unfortunately, speakers may not spontaneously use a *show* gesture with children beyond the toddler stage who are like the non-verbal ASD children in this study; however, our results indicate that they are likely to benefit.

Funding

This research was supported by funding from Ithaca College. Assistance in the collection of data was provided by the Ithaca College Research Team 04, particularly Kevin Nhui, Stephanie Miller, Ninfa Leal, Elizabeth Stuprich, Michelle Rhoades, and Elicia Wartman. Digital video clips were created at Ithaca College by David Coleman and at the School of Cinematic Arts, University of Southern California, under the direction of Greg Vannoy by Christopher Hall and Cody Charfauros.

Declaration of competing interest

None.

References

Dawson, G., Meltzoff, A.N., Osterling, J., Rinaldi, J., Brown, E., 1998. Children with autism fail to orient to naturally occurring social stimuli. J. Autism Dev. Disord. 28, 479–485.
Fenson, L., Pethick, S.J., Renda, C., Cox, J.L., Dale, P.S., Reznick, J.S., 2000. Short-form versions of the MacArthur communicative development Inventories. Appl. Psycholinguist. 21, 95–115.
Gibson, E.J., Pick, A.D., 2000. An Ecological Approach to Perceptual Learning and Development. Oxford University Press, Oxford, UK.
Gibson, J.J., 1966. The Senses Considered as Perceptual Systems. Houghton Mifflin, Boston.
Gibson, J.J., 1979. The Ecological Approach to Visual Perception. Houghton Mifflin, Boston.
Goldstein, S., Naglieri, J.A., 2010. Autism Spectrum Rating Scales (ASRS). Multi-Health Systems Inc, Toronto,Canada.
Keehn, B., Müller, R.-A., Townsend, J., 2013. Atypical attentional networks and the emergence of autism. Neurosci. Biobehav. Rev. 37, 164–183.
Laeng, B., Sirois, S., Gredebäck, G., 2012. Pupillometry: a window to the pre-conscious? Perspect. Psychol. Sci. 7, 18–27.
Rader, N., Zukow-Goldring, P., 2010. How the hands control attention during early word learning. Gesture 10, 202–221.
Rader, N., Zukow-Goldring, P., 2012. Caregivers' gestures direct infant attention during early word learning: the importance of dynamic synchrony. Lang. Sci. 34, 559–568.
Rader, N., Zukow-Goldring, P., 2015. The role of speech-gesture synchrony in clipping words from the speech stream: Evidence from infant pupil responses. Ecol. Psychol. 27, 1–10.
Read, C., Szokolszky, A., 2018. An emerging developmental ecological psychology: future directions and potentials. Ecol. Psychol. 30 (2), 174–194.
Reed, E.S., 1996. Encountering the World: Toward an Ecological Psychology. Oxford U. Press, NY.
Sahlstrom, A., Rader, N., 2009. Measuring speech-gesture synchrony in mother-to-infant interactions. In: Meeting of the Society for Research in Child Development. In preparation.
Zukow-Goldring, P., 1997. A Social Ecologial Realist Approach to the Emergence of the Lexicon: Educating Attention to Amodal Invariants in Gesture and Speech. Evolving explanations of development: ecological approaches to organism-environment systems. American Psychological Association, Washington, D.C., pp. 199–250
Zukow-Goldring, P., Rader, N., 2001. Perceiving referring actions. Dev. Sci. 4, 28–30.