

ARTICLE

# Word-object and action-object learning in a unimodal context during early childhood

Sarah Eiteljoerge<sup>1,2</sup> , Birgit Elsner<sup>3</sup> and Nivedita Mani<sup>1,2</sup>

<sup>1</sup>Psychology of Language, University of Göttingen, Göttingen, Germany

<sup>2</sup>Leibniz ScienceCampus “Primate Cognition”, Göttingen, Germany

<sup>3</sup>Developmental Psychology, University of Potsdam, Potsdam, Germany

**Corresponding author:** Sarah Eiteljoerge; Email: [sarah.eiteljoerge@psych.uni-goettingen.de](mailto:sarah.eiteljoerge@psych.uni-goettingen.de)

(Received 29 September 2023; Revised 29 January 2024; Accepted 30 January 2024)

## Abstract

Word-object and action-object learning in children aged 30 to 48 months appears to develop at a similar time scale and adheres to similar attentional constraints. However, children below 36 months show different patterns of learning word-object and action-object associations when this information is presented in a bimodal context (Eiteljoerge et al., 2019b). Here, we investigated 12- and 24-month-olds' word-object and action-object learning when this information is presented in a unimodal context. Forty 12- and 24-month-olds were presented with two novel objects that were either first associated with a novel label (word learning task) and then later with a novel action (action learning task) or vice versa. In subsequent yoked test phases, children either heard one of the novel labels or saw a hand performing one of the actions presented with the two objects on screen while we measured their target looking. Generalized linear mixed models indicate that 12-month-olds learned action-object associations but not word-object associations and 24-month-olds learned neither word- nor action-object associations. These results extend previous findings (Eiteljoerge et al., 2019b) and, together, suggest that children appear to learn action-object associations early in development while struggling with learning word-object associations in certain contexts until 2 years of age.

**Keywords:** action learning; temporal synchrony; unimodal learning; word learning

## 1. Introduction

Young children are routinely confronted with objects in their environment which are unfamiliar to them. Knowledge about these objects is typically acquired in instructional settings where informed adults tell them the names of these objects or show them the scope of actions that can be performed on them. For example, caregivers might tell the child that the animal *frog* is called a ‘frog’ and show them that it hops around. Indeed, studies suggest that children are frequently provided



with information about both the names of objects and the actions that can be performed on objects in their daily input. For instance, Gogate et al. (2000) find that between 36% and 76% of interactions with children contain synchronous temporally aligned bimodal object associations.

Given such input, infants appear to be able to learn to associate words with objects. For example, already around 6 months of age, infants show successful picture recognition upon hearing their own name, the words ‘mommy’ or ‘daddy’ or frequently heard labels for body parts (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012). Similarly, early in development, they can discriminate between matching and mismatching actions for objects (e.g., using a hairbrush to brush hair versus teeth, Bahrick et al., 2002; Hunnius & Bekkering, 2010). In other words, from a very early age, children appear to be able to learn to associate both words and actions with objects in their environment. Therefore, the mechanisms underlying such associative learning in both domains might be in place already during the first year of life. The current study examines the extent to which these mechanisms are shared across domains based on examination of the developmental trajectory of word-object and action-object association learning in early childhood.

Indeed, research has long debated the generality or specificity of the mechanisms underlying learning in these domains. On the one hand, the ease with which children learn language led researchers like Chomsky (1965) to suggest that children are born with linguistic core knowledge that can be adapted to the language of the child when provided with language-specific input (Chomsky, 1993, 2014). Thus, he suggested that language learning is facilitated by domain-specific mechanisms that are in place prenatally (see also Fodor, 1983). In contrast, other researchers highlighted the role of the environment in children’s development (Piaget, 1952; Vygotsky, 1980). Here, it is postulated that children are born with some low-level domain-general innate structures that foster learning from an early age. Knowledge acquisition, including language, is then shaped by a dynamic interaction between the environment and the child (see also Elman, 2005; Montessori, 1914; Smith & Thelen, 2003; Thelen & Smith, 1996; Tomasello, 1999). A synthesis of these accounts was suggested by Karmiloff-Smith (1992). She wrote that children might be born with biological predispositions for learning and implicit representations (e.g., numerosity, rule following) and that such representations only develop into explicit representations through the child’s interaction with the environment (e.g., counting, metalinguistic awareness). In this context, such general predispositions might include domain-general processes that may become specialized over time to adapt to the needs of a certain domain. Thus, over time, such mechanisms might turn into domain-specific mechanisms being applicable to only a certain domain.

Examining the developmental trajectory of learning in different domains can provide insights into the extent to which learning is driven by domain-general or domain-specific mechanisms. For instance, if learning in two domains (e.g., words and actions) develops in synchrony, this could support the argument that learning in these domains is driven by similar mechanisms. On the other hand, if learning does not develop in synchrony, this would either suggest that learning in these two domains is driven by different mechanisms or at the very least that it is driven by similar mechanisms which share a mediator (e.g., attention, categorization) that differently weights information from the two domains.

Studies investigating parallels in word-object and action-object association learning have shown that learning in both these domains appears to develop at a similar

time scale and adhere to similar mechanisms. For example, Childers and Tomasello (2002) presented 30-month-olds with novel nouns, novel verbs or novel actions for objects and found successful retention in comprehension tests in all conditions (see also Hahn & Gershkoff-Stowe, 2010). Childers and Tomasello (2003) further showed that 30-month-olds' extension of novel actions and nouns to novel objects followed a similar pattern across modalities. Similarly, Riggs et al. (2016) found that 3- to 4-year-olds map novel nouns and novel actions to novel objects, that children retained both types of mappings after a week and that they could extend such actions to other category members. With regard to shared mediators influencing learning across domains, Dysart et al. (2016) suggest that extending learning to other novel category members appears to be guided by novelty in both domains. Such findings have spurred suggestions that domain-general mechanisms might be at play during learning.

These studies have typically investigated children's behavior from around 30 months onward. To what extent are similarly synchronous developmental trajectories observable at younger ages? A recent study by Eiteljoerge et al. (2019b) finds that word-object and action-object association learning may follow different trajectories in early childhood. Here, 12-, 24- and 36-month-olds (as well as adults) were trained on novel actions and novel words for novel objects and tested on their recognition of the newly learned word-object and action-object associations. Twelve-month-olds learned action-object but not word-object associations, 24-month-olds learned neither action-object nor word-object associations, 36-month-olds learned word-object but not action-object associations and adults learned both types of associations.

In keeping with such differences across domains, studies on the production of novel words and actions for objects suggest an advantage for action-object associations. Childers and Tomasello (2002) showed that 30-month-olds' performance in production tests was better in action trials than in noun or verb trials. Similarly, 3- to 5-year-olds showed more correct productions of novel actions for objects relative to novel words for objects (Hahn & Gershkoff-Stowe, 2010). Together, these results might hint toward differences in the developmental trajectory across the two domains.

The inconsistencies across the studies could be explained either by the ages of the children tested or differences in the structure of the input. However, Eiteljoerge et al. (2019b) find differences across domains at 3 years of age, where Hahn and Gershkoff-Stowe (2010) find none in the comprehension task, suggesting that age alone cannot capture the differences across the studies. One possible explanation for the differences reported relates to the amount of information from multiple modalities presented concurrently across the studies.

Most of these studies have tested children in unimodal contexts (Childers & Tomasello, 2002, 2003; Dysart et al., 2016; Riggs et al., 2016). Here, one group of children was presented with novel words for objects (in the absence of any action information) and tested on their recognition of these novel word-object associations, and a separate group of children was presented with novel actions for objects and subsequently tested on their recognition of the novel action-object associations. In contrast, Eiteljoerge et al. (2019b) presented children with bimodal contexts, where they saw both the actions that could be performed on an object and heard the label for this object within the same trial, before being tested on their recognition of the novel word- and action-object associations.

As suggested by the Intersensory Redundancy Hypothesis (Bahrick et al., 2002; Bahrick & Lickliter, 2000; Gogate & Hollich, 2010), the presentation of information in unimodal and bimodal contexts could have implications for learning. Here, the presentation of information from multiple modalities in synchrony provides redundant information across senses (amodal information) which can foster learning in multimodal settings. Specifically, the presentation of synchronous multimodal information fosters learning, while the presentation of asynchronous modality-specific information in multimodal settings, as in Eiteljoerge et al. (2019b), may attenuate learning. However, the presentation of multimodal information can also lead to referential ambiguity since a label that is presented together with an action and an object could refer to either of the two. Only grammatical information would help to disambiguate (i.e., the different (sentence) structure of a noun versus a verb) in such a situation. Thus, unimodal information can provide a simplified learning environment without any ambiguity.

However, Puccini and Liszkowski (2012) found that, at 15 months of age, children learned words for objects but not the referential actions associated with these objects only when the words were presented in unimodal contexts but not when words and actions were presented together (in synchrony, but see (Werker et al., 1998)). Furthermore, the literature on modality dominance suggests a more differentiated picture of cross-domain learning, with information from different domains being prioritized, attended to and learned differently across development. Infants prioritize auditory information early on (Gogate et al., 2009); 4-year-olds switch between a visual and an auditory preference, while adults show a preference for information from the visual modality (Robinson & Sloutsky, 2004). Importantly, 4-year-olds process both visual and auditory information when information is presented separately, while maintaining an auditory preference when visual and auditory information are presented simultaneously (Sloutsky & Napolitano, 2003).

Taken together, this literature suggests that while word- and action-object association learning might follow similar trajectories (e.g., Childers & Tomasello, 2002), this might only hold for unimodal contexts where children need only to attend to information from one domain at a given time. When presented with information from multiple domains, children might prioritize information from one domain over the other.

To further evaluate the influence of bimodal and unimodal contexts on word- and action-object association learning, the current study examined the developmental trajectory of 1- and 2-year-olds' word- and action-object association learning in unimodal contexts. We presented 12- and 24-month-old children with the same novel word-object and action-object associations as in Eiteljoerge et al. (2019b), albeit now in a unimodal context. In particular, presented children with the words for an object and the actions that could be performed on an object together in the same trial, i.e., in a bimodal context. In contrast, here, in one session, children were exposed to two distinct labels for two novel objects and tested on their learning of the novel word-object associations. In a separate session, they were exposed to two distinct actions that could be performed on the same novel objects and tested on their learning of the novel action-object associations. Thus, only one modality (either a word or an action) is presented together with the object during learning, reflecting a unimodal mapping.

Three hypotheses were tested: If unimodal contexts make it easier for children to attend to information from both the word and action domain, then children in the

current study, in contrast to Eiteljoerge et al. (2019b) might learn both word- and action-object associations at the two ages tested. Alternatively, if bimodal contexts facilitate learning—at least with regard to the reported action-object association learning at 12 months of age (Eiteljoerge et al., 2019b)—both age groups might show difficulties in learning both word- and action-object associations in the unimodal context of the current study. Differences between the two age groups may also be found, given suggestions that redundant information in multimodal contexts may be especially valuable for younger children (Gogate, 2010). Thus, 12-month-olds might benefit from bimodal contexts, while 24-month-olds might be less affected by the presentation of information in unimodal contexts.

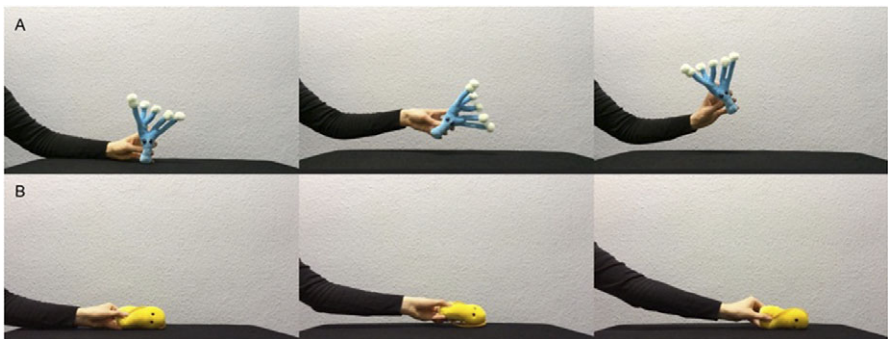
## 2. Methods

### 2.1. Participants

Forty German monolingual, typically developing 12-month-olds (mean = 11.74 months, range = 11.01–12.9 months, female = 21) and forty 24-month-olds (mean = 23.64 months, range = 21.93–25.68 months, female = 18) participated in the study. Additional participants were excluded because of preterm birth (12-month-olds: 2), visual disorder (24-month-olds: 1), unwillingness to complete the experiment (12-month-olds: 1) or because the participant did not provide sufficient data for both conditions (12-month-olds: 6; 24-month-olds: 8, see also Preprocessing). Participants were recruited via the BabyLab database, caregivers signed a written informed consent form before the study and children received a book as a reward for their participation. Ethics approval was granted by the Ethics Committee of the Georg-Elias-Müller-Institute for Psychology, University of Göttingen (Project 123).

### 2.2. Material

We used the same material as in Eiteljoerge et al. (2019b): two germs served as novel objects (<https://www.giantmicrobes.com/us/>, see Figure 1), and we used two novel labels that adhere to German phonotactic rules ('Tanu' and 'Löki') and two novel



**Figure 1.** Blue and yellow germ toys were used as novel objects. As novel actions, an upward movement with tilting to the sides (Panel a) and a sideways movement with tilting backward and forward were used (Panel b). Figure adapted from Eiteljoerge et al. (2019b).

actions ('moving up and down while tilting to the sides' and 'moving sideways with back and forward leaning').

For the training phase, we created word and action videos. In the word videos, the object was statically presented as a still frame image while it was labeled ('Schau mal, ein Tanu!' meaning 'Look, a Tanu!'). In the action videos, the object was presented in motion according to one of the actions without any linguistic information (see Figure 1). Each video was 10 s long, 720 × 420 pixels in size and presented in the middle of the screen.

For all test trials in the test phase, we used pictures of the germs which were 640 × 480 pixels in size and presented next to each other in the center of the right and left half of the screen during the baseline and the recognition phase. For the prime phase, we created recordings of a female voice asking for one of the two labels ('Wo ist das Tanu?' meaning 'Where is the Tanu?') and created videos of an empty hand moving according to one of the actions as primes (materials and data can be found on OSF at <https://doi.org/10.17605/OSF.IO/KUJPN>). This ensured that word and action prime videos were as similar as possible regarding the absence of a referent. Each of the prime videos lasted 7 s and was presented in the center of the screen.

### 2.3. Procedure

Participants sat in a dimly lit and quiet experimental room at a distance of 65 cm from a TV screen (92 × 50 cm), where the visual stimuli were presented. A remote eye tracker (Tobii X 120, Technology, 2008), set on a platform underneath the TV screen, was used to record gaze data at 60 Hz. Stimuli were presented using E-Prime software with the Tobii extension (Schneider et al., 2002). Auditory stimuli were presented through two loudspeakers above the screen. Further, two video cameras centered above the screen served to monitor the participant during the experiment. Prior to testing, gaze was calibrated using a 5-point calibration procedure, in which a red point appeared successively in every corner and also at the center of the screen. The experiment only started after successful calibration.

### 2.4. Experimental design

Each child included in the analysis participated in a word learning task and an action learning task which were administered in the same testing session but with a temporal delay between word learning and action learning sessions. For example, a child first saw the training and test phase of the word learning task, then took part in an offline administration of another task for about 30 minutes and then saw the training and test phase of the action learning task. Across children, we counterbalanced whether they received training and testing on word-object or action-object associations first.

The training phase consisted of eight trials that presented the child with the two novel objects which were either labeled (word learning task) or were moved (action learning task). Each object was presented in four trials in the center of the screen. Specific word-object and action-object associations were counterbalanced across lists, and the order of trials was randomized. Thus, 25% of the children were told that the blue object was called a Loki (in the word learning task) and saw it being moved upward and downward (in the action learning task), while the yellow object was called Tanu and saw it being moved sideways. Another 25% of the children saw



the exact opposite pairings, e.g., the yellow object was called Loki and it moved upward and downward. Another 25% of the children were told that the blue object was called a Loki (in the word learning task) and saw it being moved sideways (in the action learning task), while the yellow object was called Tanu and saw it being moved upward and downward. Another 25% saw again the exact opposite of pairings. Every child heard both labels and saw both actions. Only the mappings differed across children.

Each test phase consisted of four test trials, and each test trial consisted of a baseline phase, a prime phase and a recognition phase. The stimuli used in the test trials were identical to the stimuli used in the training phase. When the blue object from the training phase was the target of the test trial, the yellow one served as a distractor and vice versa. In the baseline phase, both objects from the training phase were presented side-by-side on-screen in silence for 2.5 s. In the word learning task, during the prime phase (7 s), both objects disappeared, and children heard the label for one of the objects while a central fixation star was on-screen. In the action learning task, children saw an empty hand performing the action for one of the objects during the same priming phase. In the following recognition phase that began 300 ms after the offset of the prime phase, in both word and action learning tasks, the two objects reappeared on-screen in the same positions for another 2.5 s.

## 2.5. Preprocessing

The eye tracker provided an estimate of where participants were looking at during each time stamp of the trial, with one data point approximately every 16 ms. All data (gaze data and trial information) were exported from E-Prime and then further processed in R (R Core Team, 2019). For each time stamp, data were only included when one or both eyes of the participant were tracked reliably (validity less than 2 on E-Prime scale). When both eyes were tracked, the mean gaze point for both eyes was computed for further analysis. Gaze data were then aggregated into 40 ms bins. Areas of interest identical in size to one another were defined based on the size and location of the videos and images on screen. For each trial in the test phase, we coded whether the participant looked at the correct object (i.e., the target), at the distractor or at neither of these on-screen.

Typically, fixation data are analyzed from 240 ms to 2000 ms from the onset of a label (see also Swingley et al., 1999), since only these fixations can be reliably construed as a response to an auditory stimulus. Following this convention, we analyzed the data from 240 ms from the onset of the baseline phase. This ensured that we avoided early fluctuations in fixations as a result of the images appearing on-screen. Furthermore, we corrected the recognition phase for any preferences in the baseline phase on an individual trial level (i.e., each time point in the recognition phase was corrected for the overall looking score in the baseline phase on that particular trial).

To ensure high data quality, single test trials were excluded if the participant looked less than  $M - 3 \times SD$  of the time to the displayed stimuli. Accordingly, trials were excluded where 12-month-olds looked less than 43% and 24-month-olds looked less than 33% to the pictures in the test trials. This led to an exclusion of 17 trials (5.15%) for 12-month-olds and 15 trials (4.46%) for 24-month-olds. Furthermore, it was important for comparing across conditions that each child contributed valid data for both conditions in the test phase. This criterion led to the exclusion of six

12-month-olds and eight 24-month-olds. Following all exclusions, a total of 80 participants with 40 in each age group were included in the final data analysis.

## 2.6. Data analysis

All analyses were done in R using the `lmer` function of the `lme4` package (Bates et al., 2014) with Gaussian error structure and identity link function. All models were checked for reasonable collinearity and absence of dispersion and influential cases, and any diversions from expected values are reported. Here, we first report generalized linear mixed models (GLMMs) that evaluate the influence of condition on proportional target looking averaged across the trial in the age groups. We then turn to GLMMs that include the factor *time* to model target looking throughout the recognition phase (also coined Growth Curve Analysis, see Mirman et al., 2008, which allows us to evaluate the influence of condition on target looking with a high temporal resolution. In all models, we included session order (i.e., word learning task followed by action learning task or action learning task followed by word learning task) as a covariate and `id`, `id` by condition, object, name and action as random effects (see Tables 1, 2 and 3 for exact model descriptions). In the models including time, we included a linear, quadratic and cubic polynomial of time in the model. Only `poly1` and `poly2` were included as random effects in the model to simplify the model structure and allow for convergence. The inclusion of three polynomials is based on theoretical expectations and practical observations, outlined in detail in Eiteljoerge et al. (2019b). In principle, if children show successful recognition of the target, target looking for young children should follow a quadratic or cubic curve if

**Table 1.** LMM testing the influence of condition on baseline-corrected proportional target looking

Age	Factor	Estimate	SE	Lower CI	Upper CI	LRT	<i>p</i>
12	Intercept	−0.002	0.036	−0.068	0.063	<sup>a</sup>	0.96
	ConditionSum	0.004	0.037	−0.065	0.072	0.011	0.918
	OrderSum	0.021	0.019	−0.016	0.055	1.223	0.269
24	Intercept	−0.01	0.032	−0.07	0.048	<sup>a</sup>	0.75
	ConditionSum	0.019	0.038	−0.056	0.092	0.239	0.625
	OrderSum	−0.001	0.026	−0.05	0.049	0.001	0.98

<sup>a</sup>Not applicable for intercept. SE: standard error; CI: confidence interval; LRT: Likelihood ratio test.

**Table 2.** GLMM evaluating the influence of condition and time on baseline-corrected proportional target looking

Age	Factor	Estimate	SE	Lower CI	Upper CI	LRT	<i>p</i>
12	Intercept	−0.006	0.041	−0.087	0.078	<sup>a</sup>	0.88
	OrderSum	0.009	0.019	−0.029	0.047	0.237	0.626
	ConditionSum:poly1	0.394	0.057	0.277	0.51	47.692	< .001
	ConditionSum:poly2	0.038	0.057	−0.073	0.146	0.445	0.505
	ConditionSum:poly3	−0.155	0.056	−0.267	−0.042	7.54	0.006
24	Intercept	−0.003	0.029	−0.06	0.052	<sup>a</sup>	0.92
	OrderSum	−0.002	0.026	−0.053	0.047	0.007	0.935
	ConditionSum:poly1	−0.057	0.06	−0.167	0.061	0.903	0.342
	ConditionSum:poly2	0.005	0.06	−0.11	0.109	0.007	0.935
	ConditionSum:poly3	0.043	0.059	−0.072	0.157	0.52	0.471

<sup>a</sup>Not applicable for intercept. SE: standard error; CI: confidence interval; LRT: Likelihood ratio test.

Significant at alpha = 0.05.



**Table 3.** GLMM evaluating the influence of time on baseline-corrected proportional target looking in the word and action condition separately

Age	Factor	Estimate	SE	Lower CI	Upper CI	LRT	<i>p</i>
12: Actions	Intercept	−0.014	0.053	−0.12	0.086	<sup>a</sup>	0.8
	poly1	0.198	0.17	−0.138	0.546	1.29	0.256
	poly2	0.04	0.147	−0.258	0.313	0.072	0.788
	poly3	−0.113	0.035	−0.183	−0.048	10.364	<b>&lt;0.001</b>
	OrderSum	0.034	0.027	−0.02	0.092	1.534	0.215
12: Words	Intercept	0.002	0.054	−0.1	0.109	<sup>a</sup>	0.97
	poly1	−0.259	0.145	−0.522	0.034	2.859	0.091
	poly2	−0.015	0.2	−0.388	0.369	0.006	0.939
	poly3	0.04	0.038	−0.038	0.111	1.101	0.294
	OrderSum	−0.011	0.022	−0.055	0.031	0.243	0.622
24: Actions	Intercept	0.001	0.05	−0.097	0.104	<sup>a</sup>	0.98
	poly1	0.097	0.248	−0.368	0.562	0.152	0.697
	poly2	0.092	0.13	−0.158	0.356	0.488	0.485
	poly3	0.014	0.038	−0.061	0.089	0.125	0.723
	OrderSum	0.002	0.036	−0.065	0.071	0.003	0.959
24: Words	Intercept	0.009	0.063	−0.122	0.131	<sup>a</sup>	0.89
	poly1	0.161	0.391	−0.595	0.878	0.167	0.683
	poly2	0.067	0.216	−0.325	0.486	0.095	0.759
	poly3	−0.016	0.035	−0.076	0.053	0.207	0.649
	OrderSum	−0.008	0.031	−0.069	0.049	0.074	0.786

<sup>a</sup>Not applicable for intercept. SE: standard error; CI: confidence interval; LRT: Likelihood ratio test. Significant at alpha = 0.05.

participants’ target looking is first at chance level looking back and forth between both objects, then increases to the target upon hearing the matching label or seeing the according action and returns to chance level when they disengage from the target stimulus (in contrast to adults who usually show a steep increase in target looking that then plateaus toward the end of the trial). A cubic term is included to reflect any looking toward the end of the trial which often stays slightly above chance level. Otherwise, if children do not show successful target recognition, target looking should stay around chance level throughout the trial.

3. Results

3.1. GLMM

For each age group, we evaluated the influence of condition on children’s proportion of fixations to the target (upon being primed by either the word or the action associated with this target) averaged across time within the trial, by running model comparisons comparing a base model without the factor condition to a model including condition. Note that the proportion of target looking during the recognition phase was baseline-corrected for any visual preferences children might show in the beginning of the test trial, i.e., during the baseline phase. Model description and results can be found in Table 1.

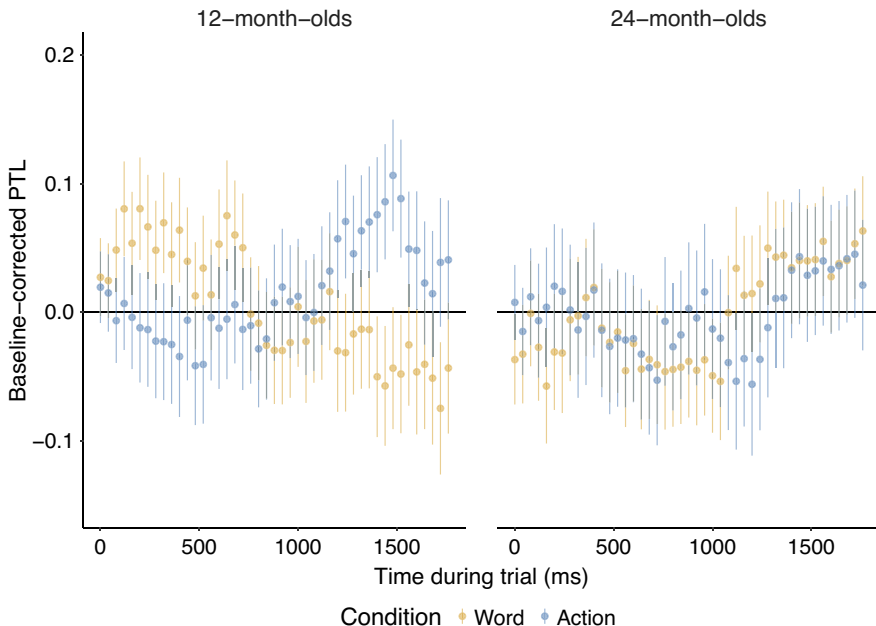
The model comparison revealed that condition did not improve model fit relative to a model without this factor at either 12 months ( $\chi^2 = 0.011$ ,  $df = 1$ ,  $p = 0.92$ ) or 24 months of age ( $\chi^2 = 0.24$ ,  $df = 1$ ,  $p = 0.63$ ). In other words, there were no differences in children’s fixations to the target when primed by either the label for this target or the action to be associated with this target when examining the total proportion of looks to the target across the entire post-priming window.

```
resbasic12<- lmer(c.PTL_corr)~ConditionSum + OrderSum +
  (1 | id) + (1 | id:Condition) + (1 | Object) + (1 | Name) + (1 | Action),
  data = subset(d_anova2, d_anova2$age == 12), REML = F, control = contr)
```

### 3.2. GLMM with temporal resolution

Next, we then added the factor Time and its linear, quadratic and cubic polynomial to the model to examine how children's target looking changed across the course of the trial in the different age groups and across conditions. For each age group, we first evaluated the influence of condition on baseline-corrected proportional target looking using model comparisons (as above) and then split the data according to condition to evaluate learning in each condition individually. The detailed results can be seen in Figure 2 and Tables 2 and 3.

For the 12-month-olds, comparing a model with the factor condition to a reduced model without condition revealed that condition contributed significantly to the model ( $\chi^2 = 58.24$ ,  $df = 4$ ,  $p < 0.001$ ). Using drop1, the model revealed a significant interaction of condition\*poly1 ( $\chi^2 = 47.69$ ,  $df = 1$ ,  $p < 0.001$ ) and condition\*poly3 ( $\chi^2 = 7.54$ ,  $df = 1$ ,  $p = 0.006$ ). These results suggest that target looking varied across time between the two conditions. We then split the data according to condition. Using drop1 revealed that there was no change in the pattern of target looking across any of the time terms in the word learning task. As is visible in Figure 2, there was a



**Figure 2.** Time course of 12- and 24-month-old baseline-corrected proportional target looking (PTL) in the word-object (yellow) and the action=object (blue) condition during the recognition phase. The line at 0 represents chance level (increase from baseline); fixations above this line indicate target looking, whereas fixations below this line indicate distractor looking. The first 240 ms are cut to allow for fixation time, and time within the trial has been corrected, so that the x-axis starts at 0.

slow but non-significant decrease in target looking throughout the trial (poly1:  $\chi^2 = 2.86$ ,  $df = 1$ ,  $p = 0.09$ ). This does not provide evidence for successful recognition of the target upon hearing its label at 12 months of age. Again, using drop1, the action-only model revealed a significant cubic change in target looking across the time course of the trial ( $\chi^2 = 10.36$ ,  $df = 1$ ,  $p = 0.001$ ). There was no effect of order in which children were exposed to word-object and action-object associations in the model, i.e., the change in target looking following priming by the actions that were associated with this target was not influenced by whether children received training in word-object or action-object associations first. Thus, we found no evidence that exposure to the objects in the first block influenced target looking across conditions when seeing these objects again in the second block.

As is visible in Figure 2, 12-month-olds' target looking in the action condition was first at chance, then increased toward the target, before it decreased again slightly toward the end of the trial.

For the 24-month-olds, comparing a model with the factor condition to a reduced model without condition revealed that condition did not contribute significantly to the model ( $\chi^2 = 1.65$ ,  $df = 4$ ,  $p = 0.8$ ). Using drop1, the model revealed no significant interactions between condition and any of the time terms. Thus, there was no evidence for differences in the pattern of children's target looking across the time course of the trial across the two conditions. For comparability with the 12-month-olds, we split the data by condition for exploratory analyses to examine the change in children's target looking in each condition separately. Neither model revealed any change in target looking across any of the time terms. As can be seen in Figure 2, the absence of any effects on the polynomials was due to target looking staying mostly around chance level with a slight increase toward the end of the trial in both the action-object and word-object conditions. There was no effect of presentation order in any of the models; in other words, children did not show any change in target looking across the time course of the trial in either condition regardless of whether they were first trained on word-object or action-object associations.

These results suggest that 12-month-olds showed increased looks to the target object when presented with the trained action association for this object, but not when presented with the trained word association for this object, suggesting that they recognized the action-object associations but not the word-object associations. 24-month-olds showed no recognition of either action-object or word-object associations in the current setting.

```
res.z.12 <- lmer(c(PTL_corr_mean) ~ ConditionSum*
(poly1 + poly2 + poly3) + OrderSum +
  (poly1 + poly2 || id) + (poly1 + poly2 | Object) + (poly1 + poly2 | Name) +
  (poly1 + poly2 |
  Action), data = subset(m.gca, m.gca$age == 12), REML = F, control = contr)
```

```
res.z.12.Words <-
lmer(c(PTL_corr_mean) ~ (poly1 + poly2 + poly3) + OrderSum +
  (poly1 + poly2 || id) + (poly1 + poly2 | Object) + (poly1 + poly2 | Name) +
  (poly1 + poly2 |
  Action), data = subset(m.gca, m.gca$age == 12 & m.gca$Condition ==
  'Word')
REML = F, control = contr)
```

#### 4. Discussion

The literature on word and action learning suggests that children appear to learn word-object and action-object associations from early on in unimodal contexts, i.e., when presented with either the labels or the actions to be associated with different novel objects (e.g., Childers & Tomasello, 2002, 2003; Dysart et al., 2016). However, they display a different pattern of learning in bimodal contexts, i.e., when presented with both the labels and actions to be associated with novel objects, showing greater attention to action-object associations early on and word-object associations later in development. Against this background, the current study re-examined children's learning of novel word-object and action-object associations in unimodal contexts. In separate eye-tracking sessions on the same day, we presented 12-month-old and 24-month-old children with novel objects that were associated either first with a novel label and then separately with a novel action (or vice versa) and tested their recognition of the trained word-object and action-object associations. The results indicate that 12-month-olds learned action-object associations when the time course of target looking was considered, but we found no evidence for learning of word-object associations in this age group. Neither did we find any evidence of learning of either word- or action-object associations at 24 months of age. This extends the findings of Eiteljoerge et al. (2019b) to word and action learning in 12- and 24-month-olds in a unimodal context.

Taken together with the results of Eiteljoerge et al. (2019b), these results highlight, on the one hand, differences in children's learning of action-object and word-object associations across development with early attention to and learning of action-object associations in both unimodal and bimodal contexts. Toward the end of the second year of life, this early focus on action information appears to recede with no evidence that children learned either word- or action-object associations. Eiteljoerge et al. (2019b) suggest that this changes further in development with 3-year-olds learning word-object associations and adults learning both. Thus, we find evidence for differences in learning of word- and action-object associations at each of the different ages tested across development.

Early in life, 12-month-olds may prioritize action information more than language information due to the greater salience of the content presented in this condition, relative to linguistic content. While language has been shown to be highly salient to young children (e.g., Fulkerson & Waxman, 2007), the changing dynamic visual content of the action presentations may be particularly attractive to young children, drawing their attention further in this condition as opposed to the word-object condition. Later in life, from around 2 years, with greater language experience and during a phase of particularly robust language development (e.g., Goldfield & Reznick, 1990), there may be changes in the prioritization of different kinds of information, with reduced attention to action and greater attention to auditory, and in particular, language input (Sloutsky & Napolitano, 2003). Indeed, Barnhart et al. (2018) suggested that this period may be associated with dynamically changing attention to different kinds of input, potentially at the individual level (c.f., Barnhart et al., 2018; Robinson & Sloutsky, 2004). In other words, it could be that some children attended more to actions at this age, while others might attend more to the language cues (Eiteljoerge et al., 2020), leading to some of them learning the action-object associations and others learning the word-object associations. Thus, children as a group at this age may not display learning of either word-object or action-object associations. Later on, with increased language proficiency, as Eiteljoerge et al. (2020)

find, children between 2 and 3 years of age then reliably learn word-object associations in the context of the task tested in the current study.

We had earlier suggested that this study could potentially also provide insights into the mechanisms underlying learning in the different domains. Indeed, the current study, together with the study by Eiteljoerge et al. (2019b), highlights developmental differences in the time course of action and word learning in young children, with 1-year-old children learning action-object associations, 2-year-olds learning neither of the associations and only older children at 3 years of age showing evidence of learning of word-object associations in the original study. This is despite children being very attentive in the training phase (looking to training stimuli was at 90% across age groups in the current study). As noted above, this may either suggest different mechanisms underlying learning in the different domains across development or that attentional constraints moderate children's learning in different domains across development. The latter possibility would also be in keeping with suggestions that early in life, children prioritize attention to certain sources of information, in this case, actions, and later, with increased perceptual abilities, find it easier to learn both word and action associations for objects simultaneously.

Indeed, there are suggestions for differential weighting of action and word information across development. For instance, Sloutsky and Napolitano (2003) suggest that children younger than 4 show auditory dominance, with greater attention to auditory as opposed to visual information early in development. Around 4, there is individual-level variation with regard to auditory and visual dominance, moving to visual dominance in adulthood (see also Barnhart et al., 2018; Robinson & Sloutsky, 2004). Alternatively, one might also suggest a change across development, with the mechanisms underlying learning early on being shared across domains (mediated by attentional constraints as above), developing with further language experience into potentially language-oriented mechanisms that highlight the connections between language and the world.

The pattern of findings reported in the current study, however, contrasts with the results of previous comprehension and production studies, albeit in older children, examining learning of novel words and novel actions for objects (Childers & Tomasello, 2003; Hahn & Gershkoff-Stowe, 2010; Riggs et al., 2016). In tasks examining children's comprehension of novel words and actions for objects, for instance, studies report no differences in learning of either kind of association in older age groups, i.e., at the age at which Eiteljoerge et al. (2019b) report improved learning of word-object associations. Indeed, if anything, at these older ages, there appear to be domain differences in production tasks, with children showing improved learning of actions for objects relative to words (Childers & Tomasello, 2002; Hahn & Gershkoff-Stowe, 2010). This disparity might be due to differences in children's attentional and perceptual abilities at the different ages tested. Above 30 months of age, children might be able to demonstrate learning and recognition of novel word-object and action-object associations in parallel in unimodal contexts. At younger ages, however, children may show different patterns of learning across domains. Indeed, Hahn and Gershkoff-Stowe (2010) report that 24-month-olds showed successful comprehension of words 59.4% of the time and of actions 35.9% of the time (with opposite results for production), similar to the more robust learning of word-object associations at later ages in Eiteljoerge et al. (2019b).

The differences in performance at 24 months between the current study and Hahn and Gershkoff-Stowe (2010) may lie in the design. In their study, children heard the label on average 8.42 times (range: 3 to 15 times). In our study, each child heard the label or saw the action for each object four times. Thus, the absence of evidence for either word or action learning in the present study might be due to the limited amount of input children received (but see Horst & Samuelson, 2008 for successful learning at 24 months after a single labeling event). Importantly, the finding that 12-month-olds showed learning of action-object associations despite this limited input (in both unimodal and bimodal contexts) highlights the early attention to action information in development.

The disparities in the findings could also be attributed to the fact that, in our design, the word or action prime was presented prior to the presentation of the target object, requiring children to maintain this prime in memory and recall it upon being presented with the target. This may have impacted how robustly children fixated the target, although we note that children at the same age show robust priming effects (Arias-Trejo & Plunkett, 2009; Mani & Plunkett, 2011). Furthermore, children were not required to disambiguate an ambiguous learning context or to retain the mapping long term which might be easier when tested in a single modality (Bion et al., 2013). Similarly, we note that, in contrast to previous studies on action-object learning (e.g., Dysart et al., 2016), we did not explicitly instruct the child to fixate the object associated with the action. Doing so in the current study would have led to an imbalance across word and action conditions, with the latter being bimodal. However, this may have led to reduced action-object learning in the 24-month-olds.

Furthermore, with regard to the comparison between Eiteljoerge et al. (2019b) and the current study, the pattern of results does not appear to differ based on whether children are presented with both action and word associations for objects together or whether they are presented with information from the two domains successively. In particular, based on previous findings that children have difficulties processing auditory and visual information when they are presented together, Eiteljoerge et al. (2019b) had suggested that the difficulties observed in 2-year-olds may have been related to the simultaneous presentation of information from both domains concurrently. The finding that 1-year-olds struggle to show word-object association learning and that 2-year-olds continue to struggle to show recognition of word-object and action-object associations in both unimodal (current study) and bimodal (Eiteljoerge et al., 2019b) contexts highlights additional constraints on learning at these ages.

Indeed, Taxitari et al. (2019) highlight similar limits to children's early word learning at 10 months of age. In that study, while 10-month-olds learned object associations for the words *cow* and *horse* (and did not show recognition of the word-object associations prior to training), the extent to which they would have been able to learn these associations without potential prior familiarity with either the objects or the labels to be trained on remains uncertain. Furthermore, children showed limited generalization of the learned labels to other members of the same category as well as limited learning from multiple exemplars of the object category. The present finding that 12-month-olds learned action-object associations but not word-object associations would, therefore, be in keeping with the constraints on early word learning suggested in Taxitari et al. (2019). While we had anticipated that learning at this younger age may have been facilitated by the simultaneous presentation of words and



action information in Eiteljoerge et al. (2019b), the finding of a very similar pattern of results in unimodal contexts in the current study suggests increased attention to and learning of action information at a young age, generally.

Note that while being presented with information from different modalities at different times did not boost learning in the current study, the Intersensory Redundancy Hypothesis (Bahrick & Lickliter, 2000) suggests that presenting children with information from both modalities in a *temporally aligned manner* boosts learning in multimodal contexts. A recent registered report finds that 2-year-olds, but not 1-year-olds, learned both word-object and action-object associations when the object was presented with both words and actions in temporal synchrony (Bothe et al., Stage2, *accepted*). Thus, presenting multimodal information with temporal alignment seems to facilitate learning for young children, compared to bimodal presentations without synchrony as in Eiteljoerge et al. (2019b) or unimodal presentations as in the current study. Indeed, the combination of the Intersensory Redundancy Hypothesis and modality dominance provides interesting avenues for new hypotheses. While the Intersensory Redundancy Hypothesis suggests that redundant information in multimodal contexts can facilitate learning, this might only hold for the dominant domain. According to Sloutsky and Napolitano (2003), children below age 4 show auditory dominance. Thus, children might learn words when temporally aligned with actions (Gogate, 2010) because words are provided as auditory information in these studies. In such a context, actions might be more difficult to learn because visual information only becomes dominant at later ages (Barnhart et al., 2018). It could also be the case that the predictions of the Intersensory Redundancy Hypothesis are suited for studies investigating cross-domain influences such as the influence of actions on word learning (Eiteljoerge et al., 2019a), but when learning in both domains is under question, the predictions might no longer hold. In such a multimodal setting, children might choose to learn only one type of information which could then be the dominant one in the scenario (Sloutsky & Napolitano, 2003).

Taken together, here, we replicate and extend earlier findings that 12-month-olds learn action-object associations but not word-object associations and that 24-month-olds learned neither action-object nor word-object associations when presented with both word and action information. Importantly, the findings remain consistent across contexts where children are presented with both word and action information concurrently as well as when this information is presented successively. The similarity in the findings between Eiteljoerge et al. (2019b) and the current study highlights, at the very least, that presenting children with multimodal contexts does not impair learning in general. Thus, across development, children show dynamic patterns of attention to information from different domains and learning of such information remains flexible across development.

**Acknowledgements.** We thank all families for participating in the study. Furthermore, we would like to thank Maurits Adam for fruitful discussions and Katharina von Zitzewitz for helping with data collection.

**Author contribution.** Sarah Eiteljoerge: Conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, roles/writing—original draft and writing—review and editing. Birgit Elsner: Conceptualization, funding acquisition, methodology, supervision and writing—review and editing. Nivedita Mani: conceptualization, formal analysis, funding acquisition, methodology, project administration, supervision, validation, visualization, roles/writing—original draft and writing—review and editing.

**Funding statement.** This research was supported by the German Research Foundation (DFG) as part of the research unit ‘Crossing the borders: The interplay of language, cognition, and the brain in early human development’ (Project: FOR 2253, grant no. EL 253/7-1), the RTG 2070 Understanding Social Relationships and the Leibniz ScienceCampus ‘Primate Cognition’.

**Competing interest.** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Arias-Trejo, N., & Plunkett, K. (2009). Lexical–semantic priming effects during infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536), 3633–3647. <https://doi.org/10.1098/rstb.2009.0146>
- Bahrnick, L. E., Gogate, L. J., & Ruiz, I. (2002). Attention and memory for faces and actions in infancy: The salience of actions over faces in dynamic events. *Child Development*, 73(6), 1629–1643. <https://doi.org/10.1111/1467-8624.00495>
- Bahrnick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, 36(2), 190–201. <https://doi.org/10.1037/0012-1649.36.2.190>
- Barnhart, W. R., Rivera, S., & Robinson, C. W. (2018). Different patterns of modality dominance across development. *Acta Psychologica*, 182, 154–165. <https://doi.org/10.1016/j.actpsy.2017.11.017>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bergelson, E., & Swingle, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253–3258. <https://doi.org/10.1073/pnas.1113380109>
- Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, 126(1), 39–53. <https://doi.org/10.1016/j.cognition.2012.08.008>
- Bothe, R., Trouillet, L., Eiteljoerge, S., Elsner, B., & Mani, N. (Stage2, accepted). *Better in sync: Temporal dynamics explain multisensory word-action-object learning in early development*.
- Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, 38(6), 967–978. <https://doi.org/10.1037/0012-1649.38.6.967>
- Childers, J. B., & Tomasello, M. (2003). Children extend both words and non-verbal actions to novel exemplars. *Developmental Science*, 6(2), 185–190. <https://doi.org/10.1111/1467-7687.00270>
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Chomsky, N. (1993). *Lectures on Government and Binding: The Pisa Lectures*. Walter de Gruyter.
- Chomsky, N. (2014). *The Minimalist Program* (20th ed.). MIT Press.
- Dysart, E. L., Mather, E., & Riggs, K. J. (2016). Young children’s referent selection is guided by novelty for both words and actions. *Journal of Experimental Child Psychology*, 146, 231–237. <https://doi.org/10.1016/j.jecp.2016.01.003>
- Eiteljoerge, S. F. V., Adam, M., Elsner, B., & Mani, N. (2019a). Consistency of co-occurring actions influences young children’s word learning. *Royal Society Open Science*, 6(8), 190097. <https://doi.org/10.1098/rsos.190097>
- Eiteljoerge, S. F. V., Adam, M., Elsner, B., & Mani, N. (2019b). Word-object and action-object association learning across early development. *PLoS One*, 14(8), e0220317. <https://doi.org/10.1371/journal.pone.0220317>
- Eiteljoerge, S. F. V., Adam, M., Elsner, B., & Mani, N. (2020). *Do preferences for words and actions influence word-object and action-object learning in early childhood?* PsyArXiv. <https://doi.org/10.31234/osf.io/jubw5>
- Elman, J. (2005). Connectionist models of cognitive development: Where next? *Trends in Cognitive Sciences*, 9(3), 111–117. <https://doi.org/10.1016/j.tics.2005.01.005>
- Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.

- Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not Tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition*, 105(1), 218–228. <https://doi.org/10.1016/j.cognition.2006.09.005>
- Gogate, L. J. (2010). Learning of syllable–object relations by preverbal infants: The role of temporal synchrony and syllable distinctiveness. *Journal of Experimental Child Psychology*, 105(3), 178–197. <https://doi.org/10.1016/j.jecp.2009.10.007>
- Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, 71(4), 878–894. <https://doi.org/10.1111/1467-8624.00197>
- Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review*, 117(2), 496–516. <https://doi.org/10.1037/a0019049>
- Gogate, L. J., Prince, C. G., & Matatyaho, D. J. (2009). Two-month-old infants' sensitivity to changes in arbitrary syllable–object pairings: The role of temporal synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 508–519. <https://doi.org/10.1037/a0013623>
- Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, 17(1), 171–183. <https://doi.org/10.1017/S0305000900013167>
- Hahn, E. R., & Gershkoff-Stowe, L. (2010). Children and adults learn actions for objects more readily than labels. *Language Learning and Development*, 6(4), 283–308. <https://doi.org/10.1080/15475441003635315>
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128–157. <https://doi.org/10.1080/15250000701795598>
- Hunniss, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, 46(2), 446–454. <https://doi.org/10.1037/a0016543>
- Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. The MIT Press.
- Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, 121(2), 196–206. <https://doi.org/10.1016/j.cognition.2011.06.013>
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475–494. <https://doi.org/10.1016/j.jml.2007.11.006>
- Montessori, M. (1914). *Dr. Montessori's own handbook*. Knopf Doubleday Publishing Group.
- Piaget, J. (1952). *The origins of intelligence in children* (Vol. 8). International Universities Press.
- Puccini, D., & Liszkowski, U. (2012). 15-month-old infants fast map words but not representational gestures of multimodal labels. *Frontiers in Psychology*, 3, 101. <https://doi.org/10.3389/fpsyg.2012.00101>
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Riggs, K. J., Mather, E., Hyde, G., & Simpson, A. (2016). Parallels between action-object mapping and word-object mapping in young children. *Cognitive Science*, 40(4), 992–1006. <https://doi.org/10.1111/cogs.12262>
- Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, 75(5), 1387–1401.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Extensions for Tobii*. Psychology Software Tools.
- Sloutsky, V. M., & Napolitano, A. C. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, 74(3), 822–833. <https://doi.org/10.1111/1467-8624.00570>
- Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, 7(8), 343–348. [https://doi.org/10.1016/S1364-6613\(03\)00156-6](https://doi.org/10.1016/S1364-6613(03)00156-6)
- Swingle, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71(2), 73–108. [https://doi.org/10.1016/S0010-0277\(99\)00021-9](https://doi.org/10.1016/S0010-0277(99)00021-9)
- Taxitari, L., Twomey, K. E., Westermann, G., & Mani, N. (2019). The limits of infants' early word learning. *Language Learning and Development*, 16(1), 1–21. <https://doi.org/10.1080/15475441.2019.1670184>
- Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT Press.
- Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, 10(2), 172–175. <https://doi.org/10.1111/1467-9280.00127>
- Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, 17(4), 432–444. <https://doi.org/10.1111/j.1532-7078.2011.00084.x>

- Tobii Technology. (2008). *User Manual: Tobii x60 and x120 eye-trackers*. Tobii.
- Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, 28, 509–529. <https://doi.org/10.1146/annurev.anthro.28.1.509>
- Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Werker, J. F., Lloyd, V. L., Cohen, L. B., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, 34(6), 1289–1309. <https://doi.org/10.1037/0012-1649.34.6.1289>

---

**Cite this article:** Eiteljoerge, S., Elsner, B., & Mani, N. (2024). Word-object and action-object learning in a unimodal context during early childhood, *Language and Cognition* 16: 1320–1337. <https://doi.org/10.1017/langcog.2024.7>