

Infants' individual sensitivity to saliency shapes early word learning in social interaction[☆]

Eugenia Wildt^{*}, Amit Singh, Katharina J. Rohlfing

Paderborn University, Faculty of Arts and Humanities, Germany

ARTICLE INFO

Keywords:

Bottom-up
Top-down
Word learning
Joint actions
Social interaction
Attention

ABSTRACT

In this paper, we conducted an extended analysis of an existing eye-tracking study (Wildt, 2023) investigating how infants' individual sensitivity to saliency shapes early word learning in social interaction. In a first phase of the study, 10-month-old infants participated in a saliency test for their attention. We used existing baseline data but applied a new analysis approach to model infants' attentional profiles—high and low preference group for salient stimuli. Afterwards, infants were exposed to the label of the referent and then participated in one of the two conditions of exploring it: In solo exploration condition (SE), infants were given the object to explore by their own; in the joint exploration condition (JE), infants could explore the object and its goal with the experimenter. In our new analysis of the test phase, we examined how differences in the sensitivity to the saliency influence infants' word learning based on the type of social interaction they receive. Our findings show that infants' looking behavior differed between JE and SE condition, with greater increases in target-directed looks over time in the JE condition. Moreover, only infants with low-preference for saliency in the JE condition successfully retrieved word-object links in all test trials. These results suggest that while joint actions can support attentional control, only infants with more balanced attention control benefit from joint actions in word learning. We conclude that early word learning is shaped by the sensitivity to the saliency in the environment, whereas social interaction helps infants to overcome the saliency bias.

1. Introduction

1.1. Away from bottom-up toward top-down processing in infancy

Infant visual attention¹ initially reflects bottom-up (exogenous) processes, driven by the sensory features of the environment. Very young infants tend to look at what is immediately visible and perceptually salient (see Wildt et al., 2019, for a review). For instance, newborns' attention is naturally drawn to stimuli with high contrast (e.g., in color or motion). Moreover, they tend to prefer the color red because their visual system is still maturing. Red, being a long-wavelength color, appears more intense and is easier to detect for their developing eyes (Brown & Lindsay, 2013). Newborns also prefer black-and-white patterns over monochromatic surfaces (Fantz, 1963), and chromatic over achromatic stimuli (Adams, 1987; Adams & Courage, 1995). Further research has revealed that five-month-olds prefer moving objects over

static stimuli (Wilcox & Clayton, 1968). This visual preference is measured by fixation duration and is referred to as the *saliency effect*, which occurs when participants fixate for longer periods on a salient stimulus compared to a less salient one (Wildt et al., 2019).

In contrast to bottom-up attention, top-down attention is primarily driven by internal goals and expectations, involving the deliberate allocation of attentional resources (Hauck et al., 2015). Early research highlights that with age, the nature of infants' attention shifts: Whereas early in infancy, looking is more influenced by stimulus properties such as physical saliency, later in infancy looking behavior reflects “top-down” factors such as infants' goals, familiarity, and the social significance of the stimuli (Colombo, 2001; Oakes, 2023). Moreover, the brain networks responsible for capturing attention develop relatively early in infancy, whereas those involved in the voluntary regulation of attention mature much later. As a result, young infants tend to focus on stimuli that are inherently attention-grabbing, and have more difficulty,

[☆] This article is part of a Special issue entitled: ‘Early Social Learning’ published in Acta Psychologica.

^{*} Corresponding author.

E-mail addresses: eugenia.wildt@upb.de (E. Wildt), amit.singh@upb.de (A. Singh), katharina.rohlfing@upb.de (K.J. Rohlfing).

¹ Visual attention—distinct from mere looking—refers to the ability to focus on selected stimuli while ignoring others in the visual environment, and it plays a critical role in how individuals learn from what they perceive (see for a review in Hunter et al., 2025).

compared to adults, in voluntarily directing their attention (Munakata et al., 2012).

However, current research argues that this view is too simplistic (Anderson, 2024), such as the development of top-down, or endogenous, attention is influenced by various factors such as the maturation of sensory and neurobiological systems, cognitive changes, and environmental inputs (Colombo & Cheatham, 2006). Theories of infant attention have undergone substantial development over the past decades (Hunter et al., 2025). Rather than assuming that infants' attention is driven solely by bottom-up or top-down mechanisms, contemporary research focuses on *how* and *when* infants show visual preferences to certain stimuli over others (e.g., Hunter and Markant, 2021; Wildt, 2023). Earlier models no longer fully capture the complexity revealed by findings from the last 25 years (Hunter et al., 2025). Current approaches instead conceptualize attention as a multifaceted system that emerges from the interaction of multiple processes (Oakes, 2023b). Crucially, infant attention is now viewed as the dynamic interplay between top-down and bottom-up control mechanisms. This conclusion makes sense when facing the memory processes involved in language learning (Wojcik, 2013): In fact, newborns are sensitive to both perceptually salient and social signals from early on. For example, they demonstrate a preference for their mother's voice over a stranger's (DeCasper & Fifer, 1980) and for faces, especially those of caregivers, over non-social stimuli (Farroni et al., 2002). Moreover, Hunter and Markant (2021) showed that 6- and 11-month-old infants focus on the most salient object in a display more often when the array contained a face of a different race, compared to a face of their own race. This indicates that infants' attention to salient objects is influenced by top-down factors, such as the familiarity of present faces. These early preferences speak to infants making use of stored representations and expectations to guide attention and learning from early on. In the following, we review current literature on individual differences in development of the dynamic interplay between both attentional systems.

1.2. Individual differences in attention profiles

The development of attention control varies across children, unfolding at different speed depending on the individual (Stevens et al., 2015). One way to investigate these individual differences is through visual behavior, which provides a window into early perception and cognition, helping us understand how infants *learn to look* and how they *look to learn* about the world around them (Oakes, 2023). Importantly, visual attention can be analyzed to derive specific parameters that act as reliable and stable indicators of individual differences in infant attention and cognitive processing.

One crucial parameter that impacts many developmental domains is the balance between bottom-up and top-down attention. Some infants predominantly rely on bottom-up attention, responding more strongly to novel or salient stimuli in their environment. Conversely, others are better able to balance bottom-up and top-down processes depending on context, utilizing their existing knowledge and expectations to guide attention toward a referent (Wass, 2022). Importantly, individual differences in attention control are linked to later socio-emotional adjustment, academic achievement (Rueda et al., 2010), and behavioral regulation (Papageorgiou et al., 2014). Studies have demonstrated that infants with stronger attentional control tend to exhibit better school performance and socio-emotional outcomes later in life, compared to peers who face more challenges in attention regulation (Papageorgiou et al., 2014; Rueda et al., 2010).

Another parameter is the length of looking times. More specifically, infants with slower-developing orienting networks ("long-lookers") exhibit prolonged fixations and difficulty shifting attention away from a stimulus, whereas infants with more mature networks ("short-lookers") switch gaze more readily. Neurologically, this is associated with maturation of the posterior orienting network, which is critical for disengaging fixation and shifting attention (Posner, 1994). Studies have

shown that short-looking infants are more likely to demonstrate evidence of recognition memory for a familiar stimulus than long-looking infants (e.g., Oakes, 2023).

Taken together, infants engage both bottom-up and top-down attentional processes from early on. But rather than a simple shift from bottom-up to top-down control, infants gradually learn to coordinate these systems. Crucially, these findings suggest that how children balance the two attentional processes in early infancy may have not only short-term implications—such as their behavior during interactions and object processing—but also long-term consequences for development across multiple domains, including language acquisition (Papageorgiou et al., 2014; Rueda et al., 2010). In everyday learning situations, infants are confronted with multiple potential referents, creating an ambiguity about what a given word might refer to. This makes the coordination of attentional processes particularly relevant for early word learning. In the following section, we therefore outline three accounts on early word learning—associative, socio-pragmatic, and interactionist—which differ in their perspective on how infants perceive salient, social, and interactional cues to learn new words.

2. Bottom-up and top-down attention in early word learning

2.1. Associationist perspective on word learning

One of the central questions in word learning is how infants learn that a word can refer to an entity (an object, an action, a property, etc.). According to the associationist perspective on word learning (e.g., Hollich et al., 2000; Moore et al., 1999; Pruden et al., 2006), infants form word meanings by associating labels with the objects or events that co-occur with them, relying on perceptual and statistical cues. In this view, infants refer a heard phonological form (a word) to the object that is most salient or temporally synchronized with naming. In essence, associationist approaches argue that infants can "break into" word learning by tuning to clear statistical regularities and object saliency without requiring rich social understanding (Hollich et al., 2000). Thus, when young infants face the challenge of identifying the referent of a word in an environment with various possibilities and distractors (Quine, 1960; Trueswell et al., 2016), the mechanism of bottom-up attention is described as particularly relevant for them, guiding them to salient and temporally contingent stimuli in their environment. However, "when infants use a perceptual strategy like this for word learning, they are also at risk for mismapping words" (Pruden et al., 2006, p. 276). Similarly, infants might overlook less salient referents or have difficulties to disengage from highly salient distractors (Frick et al., 1999). For instance, in a study by Pruden et al. (2006), 10-month-old infants were presented with two objects varying in their perceptual saliency: Depending on the condition, the experimenter labeled either the salient object (e.g., bright, colorful stimulus) or the 'boring' one. The results of the study indicate that 10-month-old infants showed a preference for learning new labels associated with highly salient objects. Moreover, infants paid less attention to the less salient object, even if it was labeled and cued with social signals. These results were interpreted to support the *Emergentist Coalition Model* (Golinkoff & Hirsh-Pasek, 2006, 2008; Pruden et al., 2006) suggesting that perceptual saliency significantly influences word learning in infants during their first year, capturing their bottom-up attention and directing it to salient referents. In this model, it is further suggested that from 12 months and older, infants begin to reweight the social and perceptual stimuli, perceiving social cues as increasingly significant source of signals. Over time, they learn the value of social cues and gradually prioritize those (eye gaze, pointing, linguistic context) as better predictors of reference (Hollich et al., 2000). These findings illustrate the developmental shift predicted by the emergentist coalition account: children begin with an associationist bias and gradually incorporate more top-down, communicative information as their experience grows. It should be critically remarked that the account does not explain what drives the change in children's

word learning. In this vein, what speaks against this account is that from early on, infants are participating in social interactions (Bruner, 1983; Rączaszek-Leonardi et al., 2013), for which they exhibit perceptual biases—for example, for familiar voices or sounds (DeCasper & Spence, 1986). It is likely that these experiences shape their language learning. Thus, in contrast to what is proposed in early stages of ECM, young infants do not seem to be solely guided by bottom-up processes but also exhibit biases toward social interaction.

2.2. Sociopragmatic perspective on word learning

Based on this social sensitivity, the **sociopragmatic perspective** views word learning as embedded in communicative exchanges, where infants interpret speakers' intentions and referential cues. Central to this account is the role of *ostensive cues* (Axelsson et al., 2012)—such as mutual gaze, infant-directed speech, or pointing—“[...] that signal to the infants that they are being addressed” (Gliga & Csibra, 2009, p. 352), indicating that an interaction is communicative and relevant for them (Csibra & Gergely, 2009). These cues function as attention organizers since they prepare infants to process incoming information as meaningful, supporting a shift from perceptual to intention-based (i.e., top-down) attention. For instance, when a caregiver looks at an object and uses infant-directed speech, infants as young as six months are more likely to follow their gaze and interpret the referent of the utterance (Senju & Csibra, 2008). Thus, ostensive cues not only attract infants' attention but also help establish referential expectations, as infants are biased to interpret such cues as indicators of communicative intent. This initiates top-down processing, which is essential for word learning. Taken together, sociopragmatic accounts highlight infants' sensitivity to communicative intent and ostensive signals, which help organize their attention toward relevant information. Ostensive cues evoke top-down attention: In ostensive contexts, infants expect the conveyed information to be relevant and trustworthy (Csibra & Gergely, 2009), thereby initiating the top-down processing essential for language acquisition.

However, evidence against this perspective comes from Szufnarowska et al. (2014), who found that six-month-old infants follow others' gaze direction in any highly attention-grabbing situation, regardless of whether these situations involve communicative intent or ostensive cues. This suggests that sensitivity to gaze following in infancy is not limited to contexts containing ostensive signals but is also elicited by attention-grabbing human actions more broadly, even in the absence of communicative intent.

2.3. Interactionist perspective on word learning

Focusing on memory processes, the **interactionist** perspective argues that the effectiveness of mere (joint) attention to the referent may be limited unless a new word is embedded within pragmatic frames (Rohlfing et al., 2016). Thus, the **interactionist perspective** (Rohlfing et al., 2016; Wildt et al., 2019) emphasizes the *structure* of interactional routines, suggesting that infants' top-down attention develops through repeated participation in socially framed action sequences. These pragmatic frames provide infants with expectations about when and how words are used, and in what contexts they become meaningful. Through structured experiences—such as routines of naming while diaper changing, playing or reading—infants learn to anticipate relevant information and allocate their attention accordingly. Wildt (2023) demonstrated this in an eye-tracking study with 10-month-old infants by putting salient, social and interactive cues in competition. Building on the interactionist perspective, we summarize the original study by Wildt (2023) as a basis for our new analysis of the existing data.

In this eye-tracking study, ten-month-old was presented with two object sets. Each object set consisted of a perceptually salient object/distractor (yellow corn cob, blue ship) and a less salient object/referent (wooden block, wooden figure) (see Wildt, 2023, Fig. 1). For the referents we used the labels *Modi*, *Doma*, *Guse*, and *Nebo*. Whereas perceptual saliency was achieved by a bright color and an interesting form, the less salient objects were dull in both color and appearance. To ensure comparability in size and shape, the corn cob and the wooden figure were paired together as one object set, and the wooden block was paired with the blue ship. Furthermore, the less salient objects had specific functions that were crucial for the training phase of the experiment in which each infant could try to achieve the functional goal of the object either alone or with support.

The study was designed on Tobii-Studio software with 120 Hz remote eye-tracker installed on the 1920 (width) pixels × 1080 (height) pixels monitor. The present study was divided into four phases followed in a chronology – exploration, saliency, training (with two conditions), and test phase (2× original-label, new-label, recovery trial). The saliency and test phases were conducted on an eye-tracking system while in the training phase, the child interacted with the object along with the experimenter. The four phases are described below (for a more detailed description see Wildt, in 2023; for eye tracking videos see in the OSF-project).

In the *exploration phase*, infants had the opportunity to explore the object set for 26 s. (see Pruden et al., 2006) in their own way (e.g., touch, put in the mouth) before seeing it on the screen.

In the *saliency phase*, infants' gaze behavior was measured with an

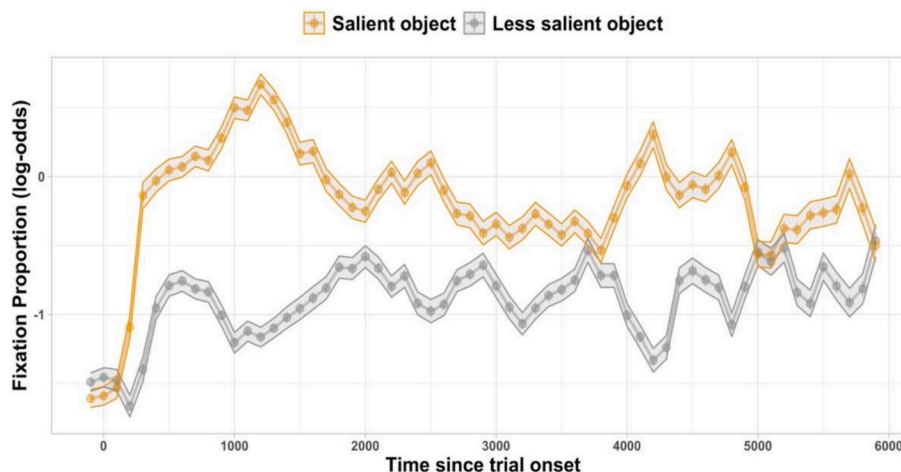


Fig. 1. Fixation proportion to the salient and less salient object as a function of time. Each time bin represents 100 ms period.

eye tracker, while they were shown two static objects for 6 s. The trial contained a salient object and a less salient object, and each infant watched two trials containing such objects placed randomly on either side of the screen. All the trials in the saliency phase were presented without labels, serving as a baseline to assess infants' bottom-up attention. Wildt (2023) demonstrated that, overall, 10-month-old infants show a saliency effect during the saliency phase, i.e., when presented with two stimuli differing in appearance, they tend to look longer at the more salient one. Thus, at baseline, young infants exhibited bottom-up attention,

The subsequent *training phase* incorporated two stages, a) the *labeling phase* and b) the *object exploration phase*. During the whole training phase, the experimenter sat across from the child while the child sat on the caregiver's lap. The less salient and the more salient object were placed in front of the infants. First, the experimenter addressed the child to ensure mutual eye contact. Then, the experimenter synchronized language and object actions to highlight the referent (by using social cues) and the word (by giving a statistical co-occurrence of the target label). The less salient target object was ostensibly labeled five times. In this phase social and salient cues competed: On the one side, infants had to segment the new word from fluent speech, and follow the experimenters' social cues to receive information about the less salient object. On the other side, the object set included a salient object that could distract infants from the less salient one. In the second stage, infants could explore the referent and received either verbal and nonverbal support from the experimenter to achieve the functional goal of the target object (Joint Exploration group, JE) or were left to freely play with the previously labeled object on their own without any support from the experimenter (Solo Exploration group, SE). This stage aimed to investigate the effect of a joint, goal-directed interaction (in the JE-group) on the word-learning process over and above the labeling and referent selection (in the SE-group). In the training phase, when salient and social cues were in conflict and accompanied by labeling, all infants relied on social information for referent selection and looked longer to the target object.

In the *testing phase*, the infants were again brought back to the eye-tracking system and presented with the same object set for 6 s while their eye movements were recorded. The test phase used four test trials to test infants' word comprehension, including two *original-label trials*, a *new-label trial*, and a *recovery trial* (see also Pruden et al., 2006). In the original-label trials infants were required to identify the previously labeled target object, while the new-label trial tested their use of the mutual exclusivity principle by introducing a completely novel label. In the recovery trial, infants were again prompted to locate the original target object. Across these trials, successful word retrieval relied on top-down attentional control, as infants needed to access stored word-object associations and apply them flexibly in response to the task demands. In the test trials, we found no evidence of word learning in either of the two groups. However, it was evident that joint actions influence infants' attention patterns by reducing the effect of perceptual saliency in the JE condition. Conversely, the attention of infants in the SE condition remained biased toward salient object during the test.

These findings suggest that infants engaged in a joint goal-directed interaction—but not in the solo exploration condition—are able to shift their attention from bottom-up to top-down processing. Wildt (2023, p. 186) explains this shift by suggesting that jointly achieving the function of a target object highlight its *relevance*, while simultaneously diminishing the perceptual saliency of competing, distracting stimuli. In pragmatic frames, infants engage in meaningful social interactions that actively shape their attentional processes. Therefore, the interactionist perspective (Rohlfing et al., 2016; Wildt, 2023; Wildt et al., 2019) underscores the contribution of social interaction and joint engagement in the word-learning process.

However, in the original analysis (Wildt, 2023), the individual variability in infants' baseline attention was not considered. This omission mirrors a common limitation also in related research (e.g., Hollich

et al., 2000; Houston-Price et al., 2006; Pruden et al., 2006) where the influence of attention on word learning is measured at population-level only to derive conclusions—leaving unexplored the ways in which individual effects affect the word learning process and the individual effects on word learning outcomes remains unexplored.

3. Extended analyses building on the study by Wildt (2023)

The present work is an extended analysis of data from an eye-tracking² study by Wildt (2023). The current analysis looks into the saliency and the testing phase. Particularly, we address the question of individual differences in bottom-up attention and how it affects the word learning in two types of interactional contexts: solo and joint exploration. We investigated this question by first estimating the group-level behavior in attention and then explicitly modeling each infant's deviance from the population estimates to identify subgroup attentional profiles.

Following this, we asked how infants with different attentional profiles and type of social interaction they received affected their word learning. Building on the interactionist perspective and the findings of Wildt (2023), it is known that infants benefit most from joint and goal-directed object exploration during social interactions. Wildt (2023) explains that jointly achieving the function of a target object serves to highlight its *relevance* for the interaction, while at the same time reducing the perceptual saliency of distracting stimuli in the environment. While we agree with this basic idea, we also build on it by taking into account the individual differences in learners' attentional profiles. We predicted that joint object explorations may be particularly beneficial for infants with a more balanced attentional profile (i.e., less dominated by bottom-up saliency) (Wass, 2022), because such infants might be making more use of social cues than infants with stronger bottom-up bias. In contrast, infants with a stronger bottom-up attentional bias—who are more reactive to salient features—may not benefit from the communicative affordances of the joint exploration. Therefore, we reasoned that effectiveness of word learning from social interactions depends not only on the interactive structure itself (joint vs. solo object exploration) but also on the infant's attentional regulation capacity.

In contrast to the first analysis (Wildt, 2023), where word learning was defined by a longer looking time to the referent in the original and recovery trial and a longer looking time to the distractor in the new-label trial, our extended analysis applied more explicit predictions for word learning. These criteria were designed to ensure that infants' looking behavior reflects retrieval of word-object mappings, rather than habituation-driven attention shifts. Specifically, we considered two eye-tracking metrics to test the word-object mapping: a) The proportion of looks toward objects across trials, and b) emergence time of such effects since the word onset. Along the provided metrics we derive following predictions for successful word learning:

a) Proportion of looks

For the gaze distribution, we predicted a significantly higher looks toward the less salient than the salient object after the onset of the target word in *original-label* and *recovery* trials. This prediction was based on the straightforward idea that if there is a correct mapping between target label and the referent object (less salient object), then a higher looks to the referent object would be observed than at the salient object, which does not match the label. Following the mutual exclusivity principle, infants should retrieve an alternative referent when a label does not match the target referent. Accordingly, for the *new-label* trial, we predicted a significant higher looks to the salient than non-salient object.

² Data files can be found on the OSF: <https://doi.org/10.17605/OSF.IO/WGXB3>.

b) Temporality of looks

Although the proportion of looks to the target and distractor provides us a reliable proxy for word–object mapping, the timing of such effect when they emerge also provides us with an additional measure of word–object retrieval link, for which eye-tracking has been suggested to have greatest strengths (Mirman, 2014). Fernald et al. (2006); Fernald and Marchman (2012) argue that spoken language comprehension unfolds in real time and therefore cannot be fully understood using only summary measures of looking time. Using eye-tracking methods, they show that children's lexical knowledge is best captured by examining both accuracy and the time course of interpretation as speech unfolds. A key finding is that between approximately 15 and 25 months, children not only become more accurate in identifying familiar words, but also significantly faster in doing so Fernald et al. (2006); Fernald and Marchman (2012). This increase in processing efficiency is important because earlier and more reliable word recognition frees cognitive resources for processing subsequent input, thereby facilitating further vocabulary learning (Fernald et al., 2006, p.113).

Whereas in previous analysis by Wildt (2023) the temporal dimension was not taken into account, here we conducted additional time-dependent analysis to specifically look into the timings of word-object mapping. Accordingly, an immediate looks to the correct referent after target word onset should reveal word learning than a delayed looks which might be due to exploration, especially when the referent is co-present against the high salient object. We therefore defined an early and late analysis windows of first 0–3000 ms and 3000–6000 ms respectively after word onset. We could reliably measure the word learning, if the effect emerges in early than late time window. Our early analysis window of 0–3000 ms after word onset is based on established standards in infant word-learning research. Bergelson and Swingley (2012) used a post-target window from 367 to 3500 ms, allowing for the minimal time needed to initiate an eye movement in response to speech and capturing the period during which infants reliably shift attention to the target. This window for infants in the first year of life is slightly longer than those used for children older than 12 months (e.g. Fernald et al., 2006), reflecting the slower processing speed of younger infants and ensuring that rapid word-object mapping is captured.

Importantly, we define successful word learning only when both criteria are met simultaneously across all eye-tracking trials: (1) a systematic pattern in the proportion of looks consistent with correct word–object mapping, and (2) a temporally early emergence of this effect within the predefined early time window. Only the convergence of both metrics provides evidence that infants are reliably retrieving learned word–object mappings, rather than exhibiting random, stimulus-driven, or exploratory looking behavior.

3.1. Participants

A total of 105 ten-month-old German-learning infants participated in this study together with one of their primary caregivers. They were from families with a middle or higher socioeconomic background as measured by their caregivers' level of education. The final sample comprised 94 infants ($M_{age} = 305$ days, $SD = 3.92$; $N_{female} = 50$, $N_{male} = 44$), as 11 participants had to be excluded from data analyses for the following reasons: poor eye-tracking data quality ($N = 4$), born before completion of the 37th week of pregnancy ($N = 4$), fussiness of the child ($N = 3$), and technical problems ($N = 1$). The final sample included 94 infants of which 49 infants belonged to the Joint Exploration (JE) group ($M_{age} = 305$ days, $SD = 3.79$; $N_{female} = 22$, $N_{male} = 27$) and 45 infants to the Solo Exploration (SE) group ($M_{age} = 304$ days, $SD = 3.48$; $N_{female} = 28$, $N_{male} = 17$).

3.2. Data analysis – pre-processing of eye-movement data

Two rectangular Areas of Interests (AOIs) were drawn each around

the salient (distractor location) and less salient (target location) on the screen. The fixations were filtered using Tobii-IV Filter inbuilt algorithm. For each trial, the overall time period was divided into 100 ms bins. To calculate the fixation proportion at each time bin, the number of fixations falling in an AOI within a bin was counted and divided by the total number of fixations falling into that bin using a R script (R Core Team, 2022). Trials with more than 25% of trackloss and infants with more than 50% of overall track loss were excluded from the analysis.

In our analysis, we visualized time-dependent changes to the fixation on the salient and less salient objects mainly for two phases of the experiment: Saliency and Testing. To calculate our dependent variable, we used the log-transformed odd ratio fixation proportion on the AOIs for each 100 ms time bins to observe the preference for the objects in each time bin (Barr, 2008). The eye-tracking data were analyzed using the Growth Curve Analysis method, GCA (Mirman, 2014; Mirman et al., 2008). As we were expected to observe the initial increase and subsequent decrease in fixation along the trial, a second-order orthogonal polynomial was fitted on the data. All the analysis were done using lme4 package (Bates et al., 2015). The training phase relied on data wherein gaze behavior was manually coded from 66 videos. Therefore, there was no need to conduct re-analyses using the video data from the study by Wildt (2023).

4. Results – eye movements

Our goal was to examine infants' fixations on the salient (distractor) versus the less salient (target) object during two phases of the experiment: the Saliency phase and the Test phase. To do this, we calculated the difference in the proportion of fixations between the salient and less salient objects, such that higher values indicate more frequent looking toward the salient object. This measure served as the dependent variable in all subsequent analyses for both phases of the experiment. We begin by presenting the results for the Saliency phase.

4.1. Saliency phase

In the saliency phase, we were interested to measure the baseline looks of each infant toward the salient with respect to the less salient object since this might provide us with the measure of the bottom-up behavior of an infant before the test phase. This was required mainly for two reasons, a) we wanted to re-establish whether there is a saliency effect, if any, present at the population level so that individual effect sizes for each infants can be estimated subsequently, and b) in Wildt (2023) we did not take temporal dimension into account to measure baseline saliency at population level which we also included in the present analysis and ruled out the aggregation effect.

To account for the variability of fixations between participants, we used a hierarchical linear model with random intercept and slope for each infant. The model incorporated the type of object (salient vs less salient) and its interaction with polynomial times as fixed effect. Additionally, we included by-subject adjustment to slope and intercept as part of model's random effects structure to account for the extent to which each infant deviates from the grand mean of looks to the salient object. The predictor variables were manually sum coded with values +1 for the salient object and – 1 for the less salient object, so that the intercept would represent the grand mean of fixation across two objects. A summary of the model is provided in Table 1 and the fixation results are shown in Fig. 1. The model suggested a main effect of object type such that the overall fixation on the salient object was significantly higher than the less salient object ($\hat{\beta} = 0.36$, 95% CI [0.16, 0.57], $p = .001$). There was also a significant interaction of the object type with the linear ($\hat{\beta} = 0.60$, 95% CI [0.55, 0.66], $p < .001$) and quadratic ($\hat{\beta} = 0.43$, 95% CI [0.37, 0.48], $p < .001$) time terms, suggesting that infants looking proportion to the salient object significantly increased with time. Importantly, the saliency effect emerged transiently in the first

Table 1

Full model estimates for the saliency phase. The orthogonal linear and quadratic time terms are coded as Time1 and Time2 respectively. “ObjectType” refers to sum contrast coded salient (+1) vs less salient (−1) object. The intercept represents the baseline which is treated as the grand mean fixation across both the objects.

Term	$\hat{\beta}$	95% CI	t	p
Intercept	−0.58	[−0.64, −0.52]	−18.17	<0.001
Time1	0.72	[0.66, 0.77]	25.07	<0.001
Time2	−0.17	[−0.22, −0.11]	−5.71	<0.001
ObjectType	0.36	[0.16, 0.57]	3.46	<0.001
Time1 × ObjectType	0.60	[0.55, 0.66]	21.18	<0.001
Time2 × ObjectType	0.43	[0.37, 0.48]	14.58	<0.001

3000 ms (early window), supporting our assumptions about early window being prominent for meaningful effects. These results together supported the previous finding in Wildt (2023) infants' preference toward salient objects, with an exception that this time we took temporal dimension into account.

4.2. Test phase

For the test phase, our aim was twofold. Firstly, we were interested in measuring the group-level effect of social interaction on word learning. Secondly, we wanted to characterize this learning in relation to each infant's visual preference toward the salient and less salient object as observed in the saliency phase. We begin our analysis by analyzing the fixation pattern on the salient and less salient object for three different trials – Original label, new label, and recovery label. For each trial, we investigated the preference toward each object across two conditions – Solo Exploration (SE) and Joint Exploration (JE). To calculate the preference toward one object against another, following the previous method in gaze analysis (Papafragou et al., 2008), we subtracted the fixation proportion of the less salient object from the salient object, such that a higher number will represent a more preference for the salient object than the less salient object. For the original-label trial, we expected a higher look toward the less salient object as a result of ‘shifting away’ from the salient object. In contrast, for the new-label trial, a correct recognition of the target referent would be manifested by an increase in the look toward the salient object, since the novel label did not correspond directly to the target (less salient) object instead it tested the use of mutual exclusivity where the alternate (salient object) was expected to attract attention. In the last trial, which is a recovery trial, a correct referent recognition should again be indexed by an increase in the looks to the less salient object from the salient object, similarly as observed in the original-label trial.

A linear mixed-effects model was fitted on the fixation proportion for the salient and less salient object for each trial type: Original label, New Label and Recovery. The model treated group (Joint- and Solo Exploration) and its interaction with polynomial time terms as the fixed effect, which was sum coded with the values, −1 for Joint Exploration and +1 for Solo Exploration following the LMM contrast coding (Schad et al., 2020). The random effects structure for the model was defined by including the random intercept and slope adjustments for each infant. The model predictions are provided in Table 2 for each trial (original, new label, and recovery) respectively. The corresponding results for each trial type are visualized in Fig. 2.

For the original-label trial, we observed a significant difference in the looking preference toward the objects between groups with respect to the linear ($\hat{\beta} = -1.41$, 95% CI [−1.67, −1.14], $p < .001$) and quadratic ($\hat{\beta} = 1.35$, 95% CI [1.08, 1.62], $p < .001$) time terms, such that JE condition showed an early preference toward the less salient object than the SE condition after the word onset. Crucially, the effect was prominent in the first 1500 ms after the target word onset (Fig. 2, top plot). This confirmed our first prediction regarding infants from the JE condition

Table 2

Full model estimates by test trials.

Term	$\hat{\beta}$	95% CI	t	p
Original-label trial				
Intercept	0.70	[0.28, 1.11]	3.29	0.001
Time1	−0.88	[−1.15, −0.62]	−6.50	<0.001
Time2	−1.84	[−2.11, −1.58]	−13.43	<0.001
Group	0.06	[−0.36, 0.48]	0.29	0.776
Time1 × Group	−1.41	[−1.67, −1.14]	−10.37	<0.001
Time2 × Group	1.35	[1.08, 1.62]	9.85	<0.001
New label trial				
Intercept	0.94	[0.36, 1.51]	3.19	0.002
Time1	−2.18	[−2.60, −1.77]	−10.39	<0.001
Time2	−1.82	[−2.23, −1.41]	−8.69	<0.001
Group	−0.02	[−0.60, 0.55]	−0.08	0.935
Time1 × Group	0.87	[0.46, 1.28]	4.15	<0.001
Time2 × Group	−1.53	[−1.94, −1.12]	−7.32	<0.001
Recovery trial				
Intercept	0.68	[0.04, 1.32]	2.10	0.039
Time1	−1.30	[−1.67, −0.92]	−6.82	<0.001
Time2	−2.65	[−3.02, −2.28]	−14.10	<0.001
Group	0.00	[−0.64, 0.64]	0.00	0.997
Time1 × Group	−2.75	[−3.12, −2.38]	−14.46	<0.001
Time2 × Group	1.61	[1.24, 1.97]	8.54	<0.001

Note. The orthogonal linear and quadratic time terms are coded as Time1 and Time2 respectively. Group refers to the sum contrast coded Solo (+1) and Joint (−1) exploration groups.

successfully retrieving the correct word–referent link in the presence of original labels.

For the new-label trial, in which an alternate label was presented, we observed a significant increase in fixation duration toward the salient object for both groups. However, as compared to the SE condition, JE condition showed early looks to the target for linear ($\hat{\beta} = 0.87$ [0.46, 1.28], 95% CI $p < .001$), and quadratic ($\hat{\beta} = -1.53$, [−1.94, −1.12] $p < .001$), time terms (Fig. 2, middle plot). These results suggest that infants from the JE condition were faster to respond to the mutual exclusivity pattern than infants from the SE condition.

In the recovery trial, we again observed a similar pattern as the original-label trial. Here, infants belonging to the JE condition showed early look to the target object as compared to the SE condition. The aim here was to re-validate the recognition pattern observed in the original-label trial and whether the infants can correctly retrieve the objects–label mapping immediately after the new-label trial. The observed preference for the salient and less-salient objects was significant with respect to for both linear ($\hat{\beta} = -2.75$, 95% CI [−3.12, −2.38], $p < .001$) and quadratic ($\hat{\beta} = 1.61$, 95% CI [1.24, 1.97], $p < .001$) time terms. As shown in Fig. 2 (bottom plot), the joint-exploration group was able to disengage their attention quickly from the salient object after the word onset, and the preference for the less-salient (target object) was maintained for the next 1000 ms afterward. In contrast, the solo-exploration group tends to show more preference toward the salient object (distractor) for first few seconds after the word onset, which only decreases in the later time window (approx. 4000–5000 ms).

Overall, these results suggested that social interaction in the JE condition facilitated infants' retrieval of the word–object link, as evidenced not only by longer looking times to the target object but also by fast, early shifts to the referent immediately after the target word.

5. Results – individual differences

5.1. Saliency phase

Our second goal was to account for the individual differences in infants' attentional profiles and model this individual variability in

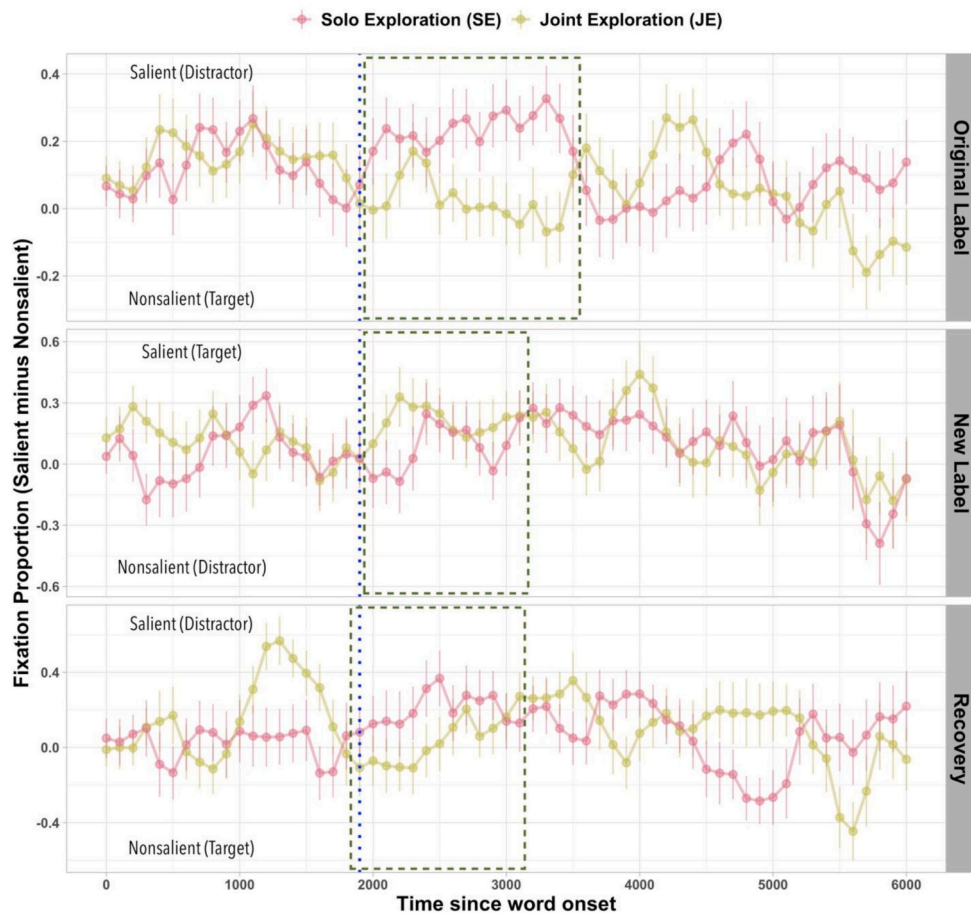


Fig. 2. Mean fixation proportion to the salient minus non-salient object for each trial type in the test phase. A high value means more preference toward the salient than the non-salient object. The vertical blue line shows the time of word onset.

saliency to investigate how this might manifest in the word learning. In addition to the type of interaction (SE vs. JE) each infant experienced, we expected the individual variability in saliency attention to also influence infants' word learning.

For this, we first extracted the effect sizes for each infant from the hierarchical model that we fitted on the data from the saliency phase (Fig. 1) and visualized the subject-level variability for the preference toward the salient object for each infant. As expected, we observed high inter-subject variability in preference toward the salient object ($SD = 1.02$), suggesting that infants' attention to the stimuli varied in terms of the degree of preference for the salient object (Fig. 3). Following this, we divided the participants into two groups based on their effect sizes: high- vs. low-preference group. A high-preference group comprised of infants having a positive effect size and likewise, a low-preference group comprised of infants having a negative effect size. This gave us a pool of infants in each group with a sample size of 40 infants in the high-preference group, and 54 infants in the low-preference group. We predicted that participants showing less preference toward the salient object in the saliency phase of the experiment would perform overall better on the word learning trials in terms of early recognition of the correct reference in response to the target labels. Moreover, in line with our results from the group-level analysis, we predicted that word learning would be facilitated when the infants engage in a joint exploration setting (JE) as compared to when they explore the object on its own

without the help of the experimenter, i.e., solo exploration (SE).

5.2. Test phase

In line with our predictions, we focused on analyzing the low- and high-preference infants who received either JE or SE conditions during the training phase. Hence, we remodeled our data after dividing the pools of infants into two groups for the test phases, a) original-label trial b) new-label trial, and c) recovery trial. The linear mixed effects model was fitted for each trial type on the test data by treating group (SE and JE) and preference (high and low) as the fixed effect structure and, by-subject adjustment for intercept, and slopes for group and preference as random effects. To remind our readers, we derived our measure of attention toward the target (less salient object) by subtracting the looks toward the target from the distractor, such that a lower fixation would now indicate a more preference toward the target. Similar to the previous analysis, the contrast coding was done as follows, group (-1 for JE; $+1$ for SE) and preference (-1 for low and $+1$ for high). The model predictions for each trial type are provided in Tables 3, 4, and 5 and the results are visualized in Figs. 4 and 5.

For the original-label trial, the model showed a significant main effect of preference ($\hat{\beta} = 0.74$, 95% CI $[0.35, 1.14]$, $p < .001$), suggesting that there was an overall higher gaze toward the salient object than the less salient (target) object. However, when infants were analyzed based

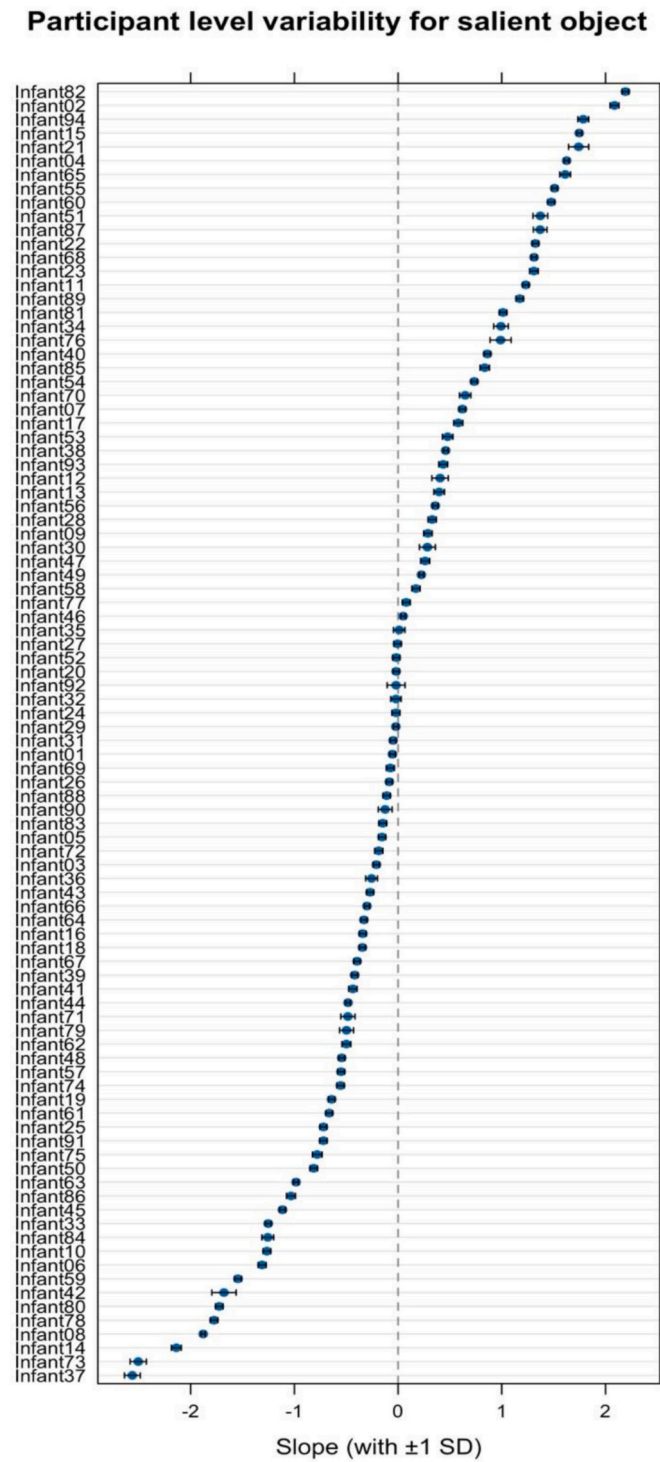


Fig. 3. Dot plot showing the participant level variability for salient and less salient object. Y-axis shows the individual participants (1–94) and the estimates are plotted against X-axis.

on their preferences style, we found that infants with low preference style gazed relatively less to the salient object (distractor) as compared to infants with high preference. This was evident by a significant interaction of preference with linear ($\hat{\beta} = -0.45, [-0.72, -0.18] p = .001$) and quadratic ($\hat{\beta} = -0.98, p < .001$) $p < .001$). Our focus was particularly on the interaction between group and preference with respect to time terms, which would tell us whether the high and low preference groups differ significantly depending upon whether they

Table 3

Full model estimates for the original-label trial by interaction group (solo and joint) and preference group (low and high). The orthogonal linear and quadratic time terms are coded as Time1 and Time2 respectively. Preference refers to sum contrast coded high and low preference group.

Term	$\hat{\beta}$	95% CI	<i>t</i>	<i>p</i>
Intercept	0.81	[0.41, 1.20]	3.97	< .001
Time1	− 0.98	[−1.25, −0.71]	− 7.10	< .001
Time2	− 2.08	[−2.36, −1.81]	− 14.92	< .001
Group	0.12	[−0.28, 0.52]	0.60	.550
Preference	0.74	[0.35, 1.14]	3.66	< .001
Time1 $\times$ Group	− 1.49	[−1.76, −1.22]	− 10.80	< .001
Time2 $\times$ Group	1.12	[0.85, 1.39]	8.01	< .001
Time1 $\times$ Preference	− 0.45	[−0.72, −0.18]	− 3.29	.001
Time2 $\times$ Preference	− 0.98	[−1.26, −0.71]	− 7.04	< .001
Group $\times$ Preference	− 0.03	[−0.43, 0.37]	− 0.16	.874
Time1 $\times$ Group $\times$ Preference	− 0.19	[−0.46, 0.08]	− 1.41	.159
Time2 $\times$ Group $\times$ Preference	− 0.92	[−1.20, −0.65]	− 6.60	< .001

Table 4

Full model estimates for the new-label trial by interaction group (solo and joint) and preference group (low and high). The orthogonal linear and quadratic time terms are coded as Time1 and Time2 respectively.

Term	$\hat{\beta}$	95% CI	<i>t</i>	<i>p</i>
Intercept	0.93	[0.35, 1.51]	3.15	0.002
Time1	−1.89	[−2.31, −1.47]	−8.86	<0.001
Time2	−1.49	[−1.91, −1.07]	−7.02	<0.001
Group	−0.02	[−0.60, 0.56]	−0.07	0.942
Preference	−0.30	[−0.87, 0.28]	−1.00	0.318
Time1 $\times$ Group	0.91	[0.50, 1.33]	4.29	<0.001
Time2 $\times$ Group	−1.46	[−1.87, −1.04]	−6.88	<0.001
Time1 $\times$ Preference	−0.76	[−1.18, −0.34]	−3.56	<0.001
Time2 $\times$ Preference	0.18	[−0.23, 0.60]	0.86	0.390
Group $\times$ Preference	0.33	[−0.25, 0.91]	1.11	0.270
Time1 $\times$ Group $\times$ Preference	2.50	[2.08, 2.92]	11.74	<0.001
Time2 $\times$ Group $\times$ Preference	1.71	[1.30, 2.13]	8.08	<0.001

Table 5

Full model estimates for the recovery-label trial by interaction group (solo and joint) and preference group (low and high). The orthogonal linear and quadratic time terms are coded as Time1 and Time2 respectively.

Term	$\hat{\beta}$	95% CI	<i>t</i>	<i>p</i>
Intercept	0.58	[−0.05, 1.21]	1.80	0.076
Time1	−1.57	[−1.95, −1.19]	−8.05	<0.001
Time2	−2.96	[−3.33, −2.58]	−15.34	<0.001
Group	−0.10	[−0.73, 0.53]	−0.32	0.751
Preference	−0.34	[−0.97, 0.29]	−1.06	0.294
Time1 $\times$ Group	−2.90	[−3.28, −2.52]	−14.90	<0.001
Time2 $\times$ Group	1.43	[1.06, 1.81]	7.44	<0.001
Time1 $\times$ Preference	−0.41	[−0.79, −0.03]	−2.09	0.036
Time2 $\times$ Preference	0.54	[0.16, 0.91]	2.78	0.005
Group $\times$ Preference	−0.53	[−1.16, 0.10]	−1.65	0.103
Time1 $\times$ Group $\times$ Preference	−0.92	[−1.30, −0.54]	−4.73	<0.001
Time2 $\times$ Group $\times$ Preference	−2.06	[−2.44, −1.68]	−10.70	<0.001

received a joint or solo treatment during the training phase. As predicted, we observed that the look to the less salient object (target) increases in response to the target word more for the JE condition than the SE condition ($\hat{\beta} = -0.92, [-1.20, -0.65], p < .001$). The effect was also evident by the reduced looks to the salient object for JE condition (Fig. 4, topright). Crucially, this pattern occurs immediately after the target word onset for the infants who belonged to low- than high-preference group. Within the high-preference group, joint exploration facilitated more looks toward the target, but this was not significantly better than the infants belonging to SE condition (Fig. 4, topleft).

Following this, we looked into the pattern specifically observed for the infants belonging to the JE condition (Fig. 5). Here, we observed

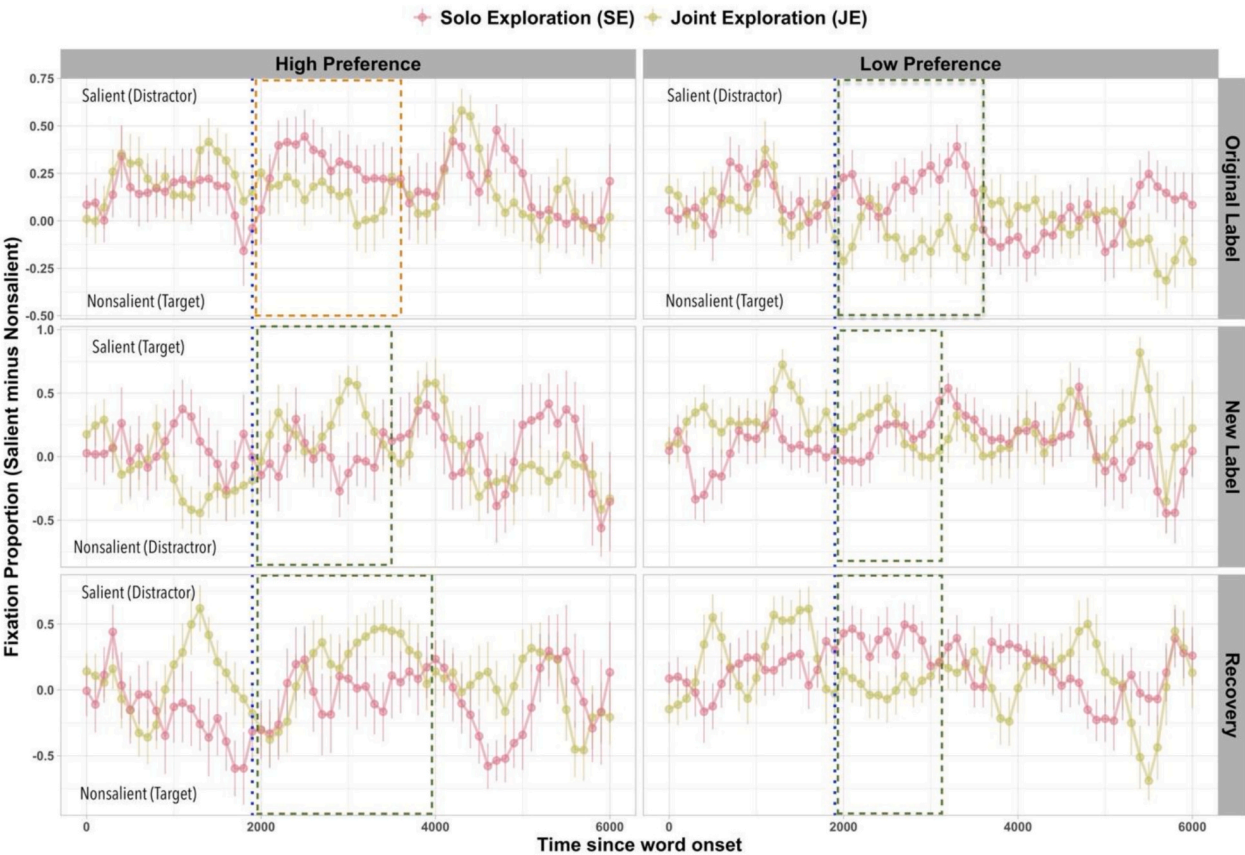


Fig. 4. The trial types are depicted in the order top to bottom: original label, new label and recovery label trials. The saliency groups are shown in columns from left to right – high preference and low preference. Increase in fixation suggests more looks toward salient than less salient object. The vertical blue line shows the time of word onset.

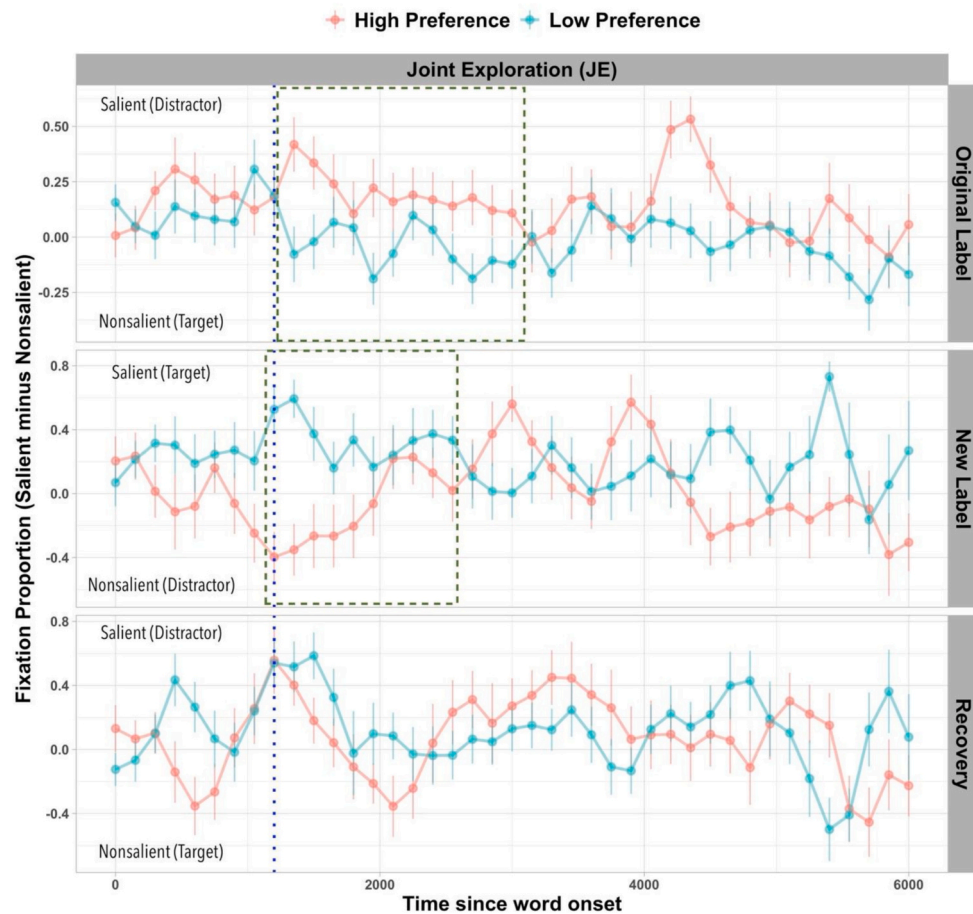


Fig. 5. Gaze behavior of all infants receiving joint exploration. The trials are shown as original label, new label and recovery. Whereas infants belonging to the low preference group were able to retrieve the word-object link, infants belonging to high preference did not.

that, whereas infants belonging to low-preference group showed high looks to the target (less-salient) in response to target word, infants belonging to high preference group kept looking more to the salient object, as shown in Fig. 5 (top plot).

Similarly, for the new-label trial, we observed a significant interaction of both time terms linear ($\hat{\beta} = 0.91$, 95% CI [0.50, 1.33], $p < .001$) and quadratic ($\hat{\beta} = -1.46$, 95% CI [-1.87, -1.04], $p < .001$) with the group (joint and solo). As can be observed in Fig. 5 (middle plot), infants belonging to JE condition looked early toward the target (salient object) as compared to the infants belonging to SE condition. We also observed that both groups (high and low preference) within the JE condition were showing visual preference toward the salient object in the first 1000 ms upon the presentation of the new label (Fig. 5 middle). However, this visual preference within the JE condition was earlier for the infants belonging to the low-preference group compared to the high-preference group (Figs. 4 and 5, middle). For the high- preference group, infants belonging to JE condition were better in identifying mutual exclusivity than infants from the SE condition (Fig. 4 middle left), as shown by higher fixation to the target (salient) after new label onset.

For the recovery trial, we observed a significant interaction of group with linear ($\hat{\beta} = -2.90$, 95% CI [-3.28, -2.52], $p < .001$) and quadratic ($\hat{\beta} = 1.43$, 95% CI [1.06, 1.81], $p < .001$) time terms, suggesting that infants overall looking preference to the objects differed across the groups (joint and solo). Here, the looks to the distractor linearly decreased with time faster for JE than SE condition ($\hat{\beta} = -2.90$).

As expected, we also observed a significant interaction of group and preference with linear ($\hat{\beta} = -0.92$, 95% CI [-1.30, -0.54], $p < .001$

and quadratic ($\hat{\beta} = -2.06$, 95% CI [-2.44, -1.68], $p < .001$) time terms, suggesting that in the early time period (as observed 2000–4000 ms on the plot), infants belonging to low-preference group and having received JE showed increased looks to the target (less-salient object) after the word onset. This pattern was opposite in the high-preference group, where even after receiving the JE, infants tend to look more toward the salient object which was the distractor (Fig. 4, bottom).

6. Discussion

Word learning in infants reflects an interplay between attention and the information provided by social interaction. In this extended analysis, we were able to pursue this interplay and further add to the previous findings of Wildt (2023) by investigating how individual differences in infants' attentional profiles influence their word learning ability. Building on attentional profiles in a baseline saliency phase of the study, we classified 10-month-olds into two groups based on their effect sizes: high- vs. low-preference group. These groups reflected infants' preferences for salient objects. Subsequently, in the training phase, after being exposed to the label for the referent object, infants explored the object either alone (SE) or with an interaction partner (JE). Thus, whereas in SE condition infants explored the object without any social help, in the JE condition they achieved the object's goal together with an interaction partner. Finally, in the test phase, we examined the effects of SE and JE on mapping abilities for infants with different attentional profiles. For our analysis, we assumed that the benefit of joint goal-directed exploration on word learning varies depending on infants' individual attentional profiles. Specifically, infants with a more balanced attentional pattern (less dominated by bottom-up saliency) were expected to show better

word learning compared to those with a stronger bottom-up preference.

We considered two eye-tracking metrics to assess successful word–object mapping: (a) the proportion of looks toward objects across trials, and (b) the emergence time of such effects after word onset. In line with our hypothesis, we found that, infants from the JE-condition overperformed infants from the SE-condition. More specifically, infants in JE-condition showed higher looking time and a fast gaze shifting to the labeled object (target) in the presence of original label, new label and recovery trials. This, at the first glance seems that infants in the JE-condition mapped objects to referents. However, when looking particularly at the interaction between group and preference, we found differences in word–object mapping between high- and low-preference groups within JE and SE conditions.

For the original-label trials, we found that infants with a low-preference attentional style were not only better at disengaging from the distractor, but also oriented more quickly to the target after word onset compared to the high-preference group. In the new-label trials, both the high- and low-preference groups showed increased attention to the salient object, which was the new target in this trial. However, the low-preference group again performed faster in the early time window. For the recovery trial, infants with a low-preference style who had received joint exploration (JE) showed increased looks to the target (the less salient object) after word onset, whereas the high-preference JE group tended to look more toward the salient object, which served as the distractor.

Taken together, our results reveal that infants with a low-preference attentional style in the joint engagement (JE) condition were the only subgroup who showed retrieval of word–referent links across all three trials. These infants consistently displayed a low preference for the salient object at baseline and during test, indicating an enhanced ability to inhibit salient distractors and to selectively drive attention to relevant stimuli. On this basis, we conclude that only this group demonstrated clear evidence of successful word learning.

Although we cannot conclude with certainty that infants in the other subgroups have not learned the word–referent links, these groups do not show consistent looking patterns across trials. For example, infants in the JE condition with a high-preference attentional style showed a pattern similar to that of the low-preference JE group in the new-label trial. However, they exhibited not only an early increase in looks toward the salient (target) object during this trial, but also a sustained preference for the salient object in the recovery trial. Based on this pattern, we suggest that infants in the high-preference JE group may have relied less on mutual exclusivity in the new-label trial than those in the low-preference JE group. Instead, their behavior appears to reflect a bottom-up attentional bias that was already present during the initial saliency phase and persisted across subsequent test trials. However, precisely disentangling whether this pattern is driven by perceptual saliency or mutual exclusivity would require a different experimental approach specifically designed to test whether bottom-up attentional guidance is the underlying mechanism.

6.1. What drives individual differences in infants' attentional profiles?

Our analysis of the saliency phase revealed significant inter-subject variability in infants' attention, with infants showing either low or high preferences for salient stimuli. These findings demonstrate that at 10 months of age, infants exhibit substantial individual variability in how they allocate attention. Specifically, we identified two distinct groups: 40 infants with a high preference for salient stimuli and 54 with a low preference. These results challenge core assumptions of the Emergentist Coalition Model (ECM; Golinkoff & Hirsh-Pasek, 2006, 2008; Pruden et al., 2006), which posits that infants in their first year primarily rely on perceptual saliency to guide attention. Rather, we observe a high variability in the perceptual saliency, which might suggest an individual propensity how infants might prioritize stimuli in the environment.

With respect to the question of what drives the individual profiles of infants with low or high preference for salient stimuli, we can only suggest some explanations. One explanation is that these differences reflect cognitive dispositions. For instance, attentional vigilance, defined as a neuromodulatory state (de Barbaro et al., 2011), can influence how infants allocate attention. Infants high in vigilance tend to focus more on salient stimuli but may have difficulties to maintain attention on less salient, but relevant, inputs. In contrast, infants with low vigilance may show better habituation and sustained attention toward less prominent stimuli (de Barbaro et al., 2011).

Furthermore, neurodevelopmental conditions such as autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) can impact attentional flexibility. Children with ASD, for example, often show less inhibition of irrelevant but salient sensory input, limiting their ability to engage with more informative stimuli (Venker et al., 2021). Similarly, children with ADHD may show reduced fixation durations and difficulties disengaging from salient distractors (Karatekin & Asarnow, 1999).

Infant's temperament also plays a role. High levels of surgency—characterized by impulsivity and high motor activity—have been associated with shorter fixation durations and faster shifts in attention during infancy (Papageorgiou et al., 2014), suggesting a temperament-based bias toward more reactive, bottom-up attention styles.

In addition to endogenous factors, another explanation is that these individual differences may emerge from environmental influences. For instance, high levels of unpredictability or 'chaos' in infants' home environment—such as noise, lack of routines, and disorganization—have been shown to reduce processing efficiency and bias infants toward bottom-up attention strategies (Tomalski et al., 2017; Wass, 2022). In such settings, children may develop top-down attentional control only later, as early on they are required to allocate greater cognitive resources to coping with unpredictability, leaving fewer resources available for the integration of goal-directed or contextual cues.

Finally, an explanation grounded in the interactionist perspective is also plausible. From early on, infants participate in social interaction, and—depending on their experience with it—they may become more or less attuned to social stimuli. In social interactions, the quality, regularity, and responsiveness of these interactions may shape their emerging sensitivity to socially relevant cues (Wildt & Rohlfing, 2024). Thus, we consider it as possible that infants' attentional profiles are shaped from birth through the experience they gained from social interactions. In this sense, attention is not merely an individual cognitive disposition but also a product of social learning, which in turn may affect how effectively infants benefit from joint action contexts in word learning.

6.2. Why does joint object manipulation impact word learning differently across attentional profiles—and how can we intervene?

Building on these group-specific findings, the concept of *pragmatic frames* (Rohlfing et al., 2016) provides a useful lens through which to interpret differences in infants' attentional profiles and their corresponding word-learning outcomes in social interaction. Repeated participation in these socially structured and contingent routines helps infants form memory traces that link the shared object or activity with linguistic input (e.g., the associated label). However, the strength and accessibility of the word–object link may vary depending on the infant's attentional profile.

Infants with a more balanced attentional profile, who show a low preference for salient properties, appear to benefit more readily from such joint object manipulations, as evidenced by their successful retrieval of the word–object link in the test. In contrast, infants whose attention is initially more guided by perceptual saliency may focus less consistently on the relevant stimuli experienced in prior interaction. Nonetheless, joint actions play a supportive role for this group by guiding their attention more toward the referent and away from

competing, salient stimuli. In our findings, however, these infants did not yet demonstrate reliable retrieval of the word–object link. This may indicate that they require a different kind or amount of support to establish a similarly robust *interaction history*.

We propose concrete interventions to investigate whether infants with a high saliency preference can benefit from tailored social learning environments. If attention profiles are shaped in part by infants' early experiences in social interaction—as we suggest from the interactionist perspective—then infants with a saliency-oriented profile may need more opportunities to engage in goal-directed joint actions to establish a robust interaction history. This additional support could take the form of *more repetitions* of object function demonstrations and labeling, to increase both the relevance of the objects and the perceptibility of new words within fluent speech. Another possibility is that, for these infants, a degree of regularity and predictability in social interaction must first be established before they can fully benefit from joint actions for word learning. In a study by Rohlfsing et al. (2018), three-year-old children diagnosed with specific language impairment (SLI) initially performed significantly worse on immediate word recall but were able to catch up with their typically developing peers in a retention test one week later—an improvement attributed to increased familiarity with the joint activity of shared book reading. Environments that are more predictable tend to foster more efficient generation and maintenance of internal attentional states through top-down control (Wass, 2022). In contrast, unpredictable environments may disrupt the development of this control, leading instead to reactive, bottom-up attention styles driven by saliency. This developmental divergence impacts how individuals engage with their surroundings: those habituated to salience-driven attention may have difficulties with goal-directed tasks requiring sustained focus. Thus, repeated and structured experiences may gradually strengthen infants' top-down attentional control and support the integration of new words into memory.

Building on these findings, we propose that for infants with a strong saliency bias, embedding learning in structured, predictable routines can foster top-down attentional control—guiding their goal-directed focus toward socially relevant cues and thus enhancing their word-learning success.

6.3. Limitations

For future research, we identify two key areas for improvement and further investigation.

a) Broader socio-demographic profiling: Despite the valuable insights gained from this study, the lack of longitudinal information about participants' profile and their development limits our understanding of how the attentional profile at 10 months of age is related to their later language development. According to previous research, the balance between bottom-up and top-down mechanisms can be influenced by various factors such as age (Bitan et al., 2009; Graziano et al., 2011; Jasso & Triesch, 2007; Moyano et al., 2022; Yurovsky & Frank, 2017), attentional vigilance (de Barbaro et al., 2011), home chaos (Evans & Wachs, 2010; Tomalski et al., 2017; Wass, 2022), (neuro)developmental disorders (Amso et al., 2014; Elison et al., 2013; Venker et al., 2021), and temperament (Papageorgiou et al., 2014). Hence, future research could incorporate a broader socio-demographic profiling and longitudinal designs to address this limitation. Moreover, more detailed demographic data, including information about the home environment—such as the level of chaos or predictability—could provide further insights into the factors influencing these attentional profiles. Additionally, our study does not provide information about the quality of the infants' social interaction experiences - for example, how responsive their parents are during interactions. Therefore, the interpretations we propose regarding the origins of individual attentional differences remain speculative. This aspect warrants further in-depth investigation to

better understand how early social interaction dynamics contribute to attentional development.

b) Optimized study design for young infants: Further limitations, e.g., regarding the study design are discussed in the previous work (Wildt, 2023). Here, we have already addressed that some infants may have failed to retrieve the word–referent link because (1) they heard the label relatively few times (only five exposures), (2) the intervening play phase had a delay that may have exceeded their retention span at 10 months, and (3) the searching task in the testing phase was not practiced before with warm up training. Future work should therefore increase label repetitions, minimize unpracticed delays, and simplify or pre-train the test task to ensure more robust word–object encoding and retrieval. Related to word–object mapping, it is worth highlighting that in the new-label trials, that is, in the mutual exclusivity condition, the target was also the salient object. This makes it difficult to clearly disentangle saliency-driven effects from top-down learning effects. Although in our experiment the JE and low-preference groups reliably showed mutual exclusivity, for the SE and high-preference groups it remains more difficult to separate learning effects from saliency effects. For further research, we also propose including additional parameters: Finally, other additional parameters (Hepach et al., 2015) such as pupil dilation (Outters et al., 2023; Wetzel et al., 2016), and saccades during word-learning tasks could provide a deeper understanding regarding infants' cognitive load and curiosity about the stimuli (e.g., Outters et al., 2023), particularly in relation to differences between infants with high and low preference styles.

In conclusion, our findings highlight that infants' individual attentional profiles, with respect to perceptual saliency, critically shape how they benefit from joint object actions in social interactions during word learning. In particular, while infants with a low preference for saliency benefit from joint engagement in learning new words, infants with a high preference for saliency seemed to require additional support. This demonstrates the need to develop tailored support strategies—such as increased repetition, predictable routines, or enriched labeling—that specifically take into account infants with different attentional capacities.

CRedit authorship contribution statement

Eugenia Wildt: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Amit Singh:** Writing – review & editing, Formal analysis. **Katharina J. Rohlfsing:** Writing – review & editing, Supervision.

Declaration of competing interest

We hereby confirm that the submitted manuscript is the result of our original work and that all authors have contributed significantly to its conception, design, analysis, and writing. The research was conducted independently, and no unauthorized external contributions have been used in the preparation of the manuscript.

All sources and contributions of others have been properly acknowledged and cited. The manuscript has not been published previously and is not under consideration for publication elsewhere.

We affirm that there are no undisclosed conflicts of interest that could be perceived as influencing the research.

All authors have read and approved the final version of the manuscript and agree to its submission to *Acta Psychologica*.

Data availability

Data files can be found in OSF-Project (<https://doi.org/10.17605/OSF.IO/WGXB3>):https://osf.io/wgxb3/overview?view_only=07b0a9152f8a418aa7aecff297cc03d2.

References

- Adams, R. J. (1987). An evaluation of color preference in early infancy. *Infant Behavior and Development*, 10(2), 143–150. [https://doi.org/10.1016/0163-6383\(87\)90029-4](https://doi.org/10.1016/0163-6383(87)90029-4)
- Adams, R. J., & Courage, M. L. (1995). Development of chromatic discrimination in early infancy. *Behavioural Brain Research*, 67(1), 99–101. [https://doi.org/10.1016/0166-4328\(94\)00143-4](https://doi.org/10.1016/0166-4328(94)00143-4)
- Amso, D., Haas, S., Tenenbaum, E., Markant, J., & Sheinkopf, S. J. (2014). Bottom-up attention orienting in young children with autism. *Journal of Autism and Developmental Disorders*, 44(3), 664–673. <https://doi.org/10.1007/s10803-013-1925-5>
- Anderson, B. A. (2024). Trichotomy revisited: A monolithic theory of attentional control. *Vision Research*, 217, Article 108366.
- Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00088>
- de Barbaro, K., Chiba, A., & Deák, G. O. (2011). Micro-analysis of infant looking in a naturalistic social setting: Insights from biologically based models of attention: Micro-analysis of infant looking. *Developmental Science*, 14(5), 1150–1160. <https://doi.org/10.1111/j.1467-7687.2011.01066.x>
- Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474. <https://doi.org/10.1016/j.jml.2007.09.002>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). 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>
- Bitan, T., Cheon, J., Lu, D., Burman, D. D., & Booth, J. R. (2009). Developmental increase in top-down and bottom-up processing in a phonological task: An effective connectivity, fMRI study. *Journal of Cognitive Neuroscience*, 21(6), 1135–1145. <https://doi.org/10.1162/jocn.2009.21065>
- Brown, A. M., & Lindsay, D. T. (2013). Infant color vision and color preferences: A tribute to Davida Teller. *Visual Neuroscience*, 30(5–6), 243–250. <https://doi.org/10.1017/S0952523813000114>
- Bruner, J. S. (1983). *Child's talk: Learning to use language*. Oxford University Press.
- Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, 52(1), 337–367. <https://doi.org/10.1146/annurev.psych.52.1.337>
- Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. *Advances in Child Development and Behavior*, 34, 283–322. [https://doi.org/10.1016/S0065-2407\(06\)80010-8](https://doi.org/10.1016/S0065-2407(06)80010-8)
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208(4448), 1174–1176. <https://doi.org/10.1126/science.7375928>
- DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, 9(2), 133–150. [https://doi.org/10.1016/0163-6383\(86\)90025-1](https://doi.org/10.1016/0163-6383(86)90025-1)
- Elison, J. T., Paterson, S. J., Wolff, J. J., Reznick, J. S., Sasson, N. J., Gu, H., ... for the IBIS Network. (2013). White matter microstructure and atypical visual orienting in 7-month-olds at risk for autism. *American Journal of Psychiatry*, 170(8), 899–908. <https://doi.org/10.1176/appi.ajp.2012.12091150>
- Evans, G. W., & Wachs, T. D. (2010). *Chaos and its influence on children's development*. 6 pp. 66–80. Washington, DC: American Psychological Association (2–3).
- Fant, R. L. (1963). Pattern vision in newborn infants. *Science*, 140(3564), 296–297. <https://doi.org/10.1126/science.140.3564.296>
- Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, 99(14), 9602–9605. <https://doi.org/10.1073/pnas.152159999>
- Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, 83(1), 203–222. <https://doi.org/10.1111/j.1467-8624.2011.01692.x>
- Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, 42(1), 98. <https://doi.org/10.1037/0012-1649.42.1.98>
- Frick, J. E., Colombo, J., & Saxon, T. F. (1999). Individual and developmental differences in disengagement of fixation in early infancy. *Child Development*, 70(3), 537–548. JSTOR.
- Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science*, 20(3), 347–353. <https://doi.org/10.1111/j.1467-9280.2009.02295.x>
- Golinkoff, R. M., & Hirsh-Pasek, K. A. (2006). Baby wordsmith: From associationist to social sophisticated. *Current Directions in Psychological Science*, 15(1), 30–33. <https://doi.org/10.1111/j.0963-7214.2006.00401.x>
- Golinkoff, R. M., & Hirsh-Pasek, K. A. (2008). How toddlers begin to learn verbs. *Trends in Cognitive Sciences*, 12(10), 397–403. <https://doi.org/10.1016/j.tics.2008.07.003>
- Graziano, P. A., Calkins, S. D., & Keane, S. P. (2011). Sustained attention development during the toddlerhood to preschool period: Associations with toddlers' emotion regulation strategies and maternal behaviour. *Infant and Child Development*, 20(6), 389–408. <https://doi.org/10.1002/icd.731>
- Hauck, M., Domnick, C., Lorenz, J., Gerloff, C., & Engel, A. K. (2015). Top-down and bottom-up modulation of pain-induced oscillations. *Frontiers in Human Neuroscience*, 9. <https://doi.org/10.3389/fnhum.2015.00375>
- Hepach, R., Vaish, A., & Tomasello, M. (2015). Novel paradigms to measure variability of behavior in early childhood: Posture, gaze, and pupil dilation. *Frontiers in Psychology*, 6, 858. <https://doi.org/10.3389/fpsyg.2015.00858>
- Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., ... Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, i–135.
- Houston-Price, C., Plunkett, K., & Duffy, H. (2006). The use of social and salience cues in early word learning. *Journal of Experimental Child Psychology*, 95(1), 27–55. <https://doi.org/10.1016/j.jecp.2006.03.006>
- Hunter, B. K., & Markant, J. (2021). Differential sensitivity to species- and race-based information in the development of attention orienting and attention holding face biases in infancy. *Developmental Psychobiology*, 63(3), 461–469. <https://doi.org/10.1002/dev.22027>
- Hunter, B. K., Kizildere, E., Klotz, S. M., Nelson, C. M., Markant, J., & Oakes, L. M. (2025). What have we learned about infant visual attention in the first 25 years of the 21st century? *Infant Behavior & Development*, 81, Article 102135. <https://doi.org/10.1016/j.infbeh.2025.102135>
- Jasso, H., & Triesch, J. (2007). Learning to attend – From bottom-up to top-down. In L. Paletta, & E. Rome (Eds.), vol. 4840. *Attention in cognitive systems. Theories and systems from an interdisciplinary viewpoint*. WAPCV 2007. Lecture notes in computer science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-77343-6_7
- Karatekin, C., & Asarnow, R. F. (1999). Exploratory eye movements to pictures in childhood-onset schizophrenia and attention-deficit/hyperactivity disorder (ADHD). *Journal of Abnormal Child Psychology*, 27, 35–49. <https://doi.org/10.1023/A:1022662323823>
- Mirman, D. (2014). Growth curve analysis: A hands-on tutorial on using multilevel regression to analyze time course data. In, Vol. 36. *Proceedings of the Annual Meeting of the Cognitive Science Society* (No. 36).
- 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>
- Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology*, 35(1), 60.
- Moyano, S., Conejero, Á., Fernández, M., Serrano, F., & Rueda, M. R. (2022). Development of visual attention control in early childhood: Associations with temperament and home environment. *Frontiers in Psychology*, 13, Article 1069478. <https://doi.org/10.3389/fpsyg.2022.1069478>
- Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing cognitive control: Three key transitions. *Current Directions in Psychological Science*, 21(2), 71–77. <https://doi.org/10.1177/0963721412436807>
- Oakes, L. M. (2023). The cascading development of visual attention in infancy: Learning to look and looking to learn. *Current Directions in Psychological Science*, 32(5), 410–417. <https://doi.org/10.1177/09637214231178744>
- Outters, V., Hepach, R., Behne, T., & Mani, N. (2023). Children's affective involvement in early word learning. *Scientific Reports*, 13(1), 7351. <https://doi.org/10.1038/s41598-023-34049-3>
- Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108(1), 155–184. <https://doi.org/10.1016/j.cognition.2008.02.007>
- Papageorgiou, K. A., Smith, T. J., Wu, R., Johnson, M. H., Kirkham, N. Z., & Ronald, A. (2014). Individual differences in infant fixation duration relate to attention and behavioral control in childhood. *Psychological Science*, 25(7), 1371–1379. <https://doi.org/10.1177/0956797614531295>
- Posner, M. I. (1994). Attention: The mechanisms of consciousness. *Proceedings of the National Academy of Sciences*, 91(16), 7398–7403. <https://doi.org/10.1073/pnas.91.16.7398>
- Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development*, 77(2), 266–280. <https://doi.org/10.1111/j.1467-8624.2006.00869.x>
- Quine, W. V. O. (1960). *Word and object*. MIT Press.
- R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing (<https://www.R-project.org>).
- Rączaszk-Leonardi, J., Nomikou, I., & Rohlfing, K. J. (2013). Young children's dialogical actions: The beginnings of purposeful intersubjectivity. *IEEE Transactions on Autonomous Mental Development*, 5(3), 210–221. <https://doi.org/10.1109/TAMD.2013.2273258>
- Rohlfing, K. J., Ceuremans, J., & Horst, J. S. (2018). Benefits of repeated book readings in children with SLI. *Communication Disorders Quarterly*, 39(2), 367–370. <https://doi.org/10.1177/1525740117692480>
- Rohlfing, K. J., Wrede, B., Vollmer, A. L., & Oudeyer, P. Y. (2016). An alternative to mapping a word onto a concept in language acquisition: Pragmatic frames. *Frontiers in Psychology*, 7, 470. <https://doi.org/10.3389/fpsyg.2016.00470>
- Rueda, M. R., Checa, P., & Rothbart, M. K. (2010). Contributions of attentional control to socioemotional and academic development. *Early Education & Development*, 21(5), 744–764. <https://doi.org/10.1080/10409289.2010.510055>
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, Article 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, 18(9), 668–671.
- Stevens, C., Paulsen, D., Yassen, A., & Neville, H. (2015). Atypical auditory refractory periods in children from lower socio-economic status backgrounds: ERP evidence for a role of selective attention. *International Journal of Psychophysiology*, 95(2), 156–166. <https://doi.org/10.1016/j.ijpsycho.2014.06.017>

- Szufnarowska, J., Rohlfling, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention? *Scientific Reports*, 4(1). <https://doi.org/10.1038/srep05304>
- Tomalski, P., Marczuk, K., Pisula, E., Malinowska, A., Kawa, R., & Niedźwiecka, A. (2017). Chaotic home environment is associated with reduced infant processing speed under high task demands. *Infant Behavior and Development*, 48, 124–133. <https://doi.org/10.1016/j.infbeh.2017.04.007>
- Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, 148, 117–135. <https://doi.org/10.1016/j.cognition.2015.11.002>
- Venker, C. E., Mathée, J., Neumann, D., Edwards, J., Saffran, J., & Ellis Weismer, S. (2021). Competing perceptual salience in a visual word recognition task differentially affects children with and without autism spectrum disorder. *Autism Research*, 14(6), 1147–1162. <https://doi.org/10.1002/aur.2457>
- Wass, S. (2022). *Contingencies and periodicities: Two types of predictability in children's real-world environment that might influence the executive control of attention [Preprint]*. PsyArXiv. <https://doi.org/10.31234/osf.io/z6rjc>
- Wetzel, N., Buttelmann, D., Schieler, A., & Widmann, A. (2016). Infant and adult pupil dilation in response to unexpected sounds. *Developmental Psychobiology*, 58(3), 382–392. <https://doi.org/10.1002/dev.21377>
- Wilcox, B. M., & Clayton, F. L. (1968). Infant visual fixation on motion pictures of the human face. *Journal of Experimental Child Psychology*, 6(1), 22–32. [https://doi.org/10.1016/0022-0965\(68\)90068-4](https://doi.org/10.1016/0022-0965(68)90068-4)
- Wildt, E. (2023). Joint actions affect infants' cognitive word–object processing. *Language, Interaction and Acquisition*, 14(2), 167–190. <https://doi.org/10.1075/lia.21024.wil>
- Wildt, E., & Rohlfling, K. J. (2024). The impact of maternal gaze responsiveness on infants' gaze following and later vocabulary development. *Infant Behavior and Development*, 74, Article 101917. <https://doi.org/10.1016/j.infbeh.2023.101917>
- Wildt, E., Rohlfling, K. J., & Scharlau, I. (2019). The role of saliency in learning first words. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01150>
- Wojcik, E. H. (2013). Remembering new words: Integrating early memory development into word learning. *Frontiers in Psychology*, 4, 151. <https://doi.org/10.3389/fpsyg.2013.00151>
- Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science*, 20(2), Article e12349. <https://doi.org/10.1111/desc.12349>