# Investigating the Effects of Multimodal Behaviors on Eliciting Joint Attention in Infants

Naomi Esparza

Committee: Dr. Gedeon Deák, Dr. Shannon Ellis, Yueyan Tang

University of California, San Diego

Cognitive Science Honors Program, 2024 - 2025

**Abstract**

Joint attention episodes are important in early learning (Tomasello, 1986) as they can help older infants and toddlers associate an adult's attentional focus with informative extrinsic events or with social input such as unfamiliar words (Baldwin, 1991; Deák and Tang, in press). Caregivers produce a variety of behavioral cues to attempt to elicit shared attention with their infants (Bard et al., 2021; Deák et al, 2014, 2018; Tang & Deák, 2024). In this paper we investigate (in the dataset described by Tang et al., 2023) how different combinations and sequences of attention-directing cues (gaze, pointing, and speech) influence the effectiveness of mothers' attention bids to younger infants – i.e., from 6 to 9 months of age – in a naturalistic setting. A sample of 48 infant-mother dyads from a Southern California city were videotaped every month from 6 to 9 months of age. Mothers were middle-class, English-fluent, and relatively well-educated. Each session included two phases: a free play phase, followed by a semi-structured (e.g., predetermined targets and target locations) but unscripted attention-sharing phase. In this joint attention phase (the focus of the current study), mothers were instructed to draw their infants' attention to three target puppets. Findings indicated a statistically significant positive correlation between the number of cues within bids and hit rates, and suggested the addition of verbal cues to be associated with higher hit rates.

## 1.1 Introduction

Before learning to speak, infants learn to coordinate visual attention with social partners as a form of communication. This coordination is a form of Joint Attention (JA): the shared attention to a single event or object by multiple individuals. Engaging in shared attention is potentially important for multiple aspects of infant development, including language development (Bruinsma et. al., 2004) and learning (Striano et. al., 2006). Through recurrent interactive episodes between adults and infants, infants might be able to make connections between the adult's attentional focus and their linguistic output. This may help infants learn the intended referents of words, thereby developing receptive vocabulary, a core faculty for communicative and linguistic skills (Tomasello et al., 1986).

In (Deák and Tang, 2024), multimodal behaviors produced by caregivers during play sessions were zoned in on to observe some influences on Joint Attention Episodes. Some commonly observed multi modal behaviors include gaze, point, verbal output, object handling, etc., all of which can facilitate and scaffold attention-following events. This study explored the dynamics of JA in naturalistic infant-caregiver interactions, and it was observed that between 6 and 9 months, gaze-led bids and the incorporation of both gaze and point within a bid made up the majority of cues used by mothers to initiate shared attention. In addition, it was found that the use of only gaze or point within a bid was rare, and that most bids included multiple cues.

In contrast to Deák and Tang's results, (Kaplan and Yu, 2024) conducted analysis on multimodal pathways to JA in naturalistic contexts, and found that for both child-led and parent-led JA bouts, the party that is not initiating JA were most likely to follow their partner's

hand (includes points), while parents were also most likely to do so in combination with following the child's gaze. In addition, almost all parents were found to participate in gaze-following in conjunction with hand following, and consistent with Deák and Tang's findings, very few participants exclusively used only gaze or point.

In (Deák et. al, 2008) the study showed eliciting verbalizations increased the infant's looks to their parents, pointing gestures were equally effective as combining directive verbalizations with eliciting verbalizations, and that directive verbalizations were more effective in attention directing than eliciting verbalizations. (will define DV and EV) The same study also determined that the combination of gaze + verbalization and gaze + point produced the highest proportion of infants who followed the care giver's cue to the correct target.

Many studies have analyzed relationships between individuals as well as limited combinations of multimodal behaviors and Joint attention between mothers and their infants. Yet few studies have analyzed the relationship between differing sequences of maternal multimodal behaviors, and initiating shared attention with an infant. Past studies have shown that the exclusive use of pointing gestures or gazes were rare, and that the hit rate differences between the two were not statistically significant (Deák and Tang, 2024). Association between verbal cues and attention following have also been studied, yet exclusive use of verbalizations were found to be no more effective than only pointing. (Deák et. al, 2008) In addition, past research has found that those who initiate shared attention are more likely to use a variety of multimodal behaviors. Considering these factors, determining the sequence of these behaviors, or cues, could lead to new findings regarding predictors for JA episodes. An infant's ability to share attention with caregivers contributes to social learning and communication (Tang, Gonzalez, Deák, 2023), and has been found to be a predictor for early language development, vocabulary size, ASD (Autism Spectrum Disorder), and has special importance for acquiring new language. (Tomasello, 1986) Therefore, understanding predictors for shared attention in mother-infant dyads could help educators develop effective methods for language development and acquisition, as well as help researchers develop improved methods for early ASD recognition and intervention .

In this study, we aim to analyze the relationship between sequences of multimodal behaviors from maternal figures when initiating shared attention with their infant. The multimodal behaviors that we will be focusing on will be: point, gaze, and verbal. We will be focusing on 6-9 month mother-infant dyads during play sessions and explore the dynamic of JA and multimodal behavior sequences. Through this analysis we will be able to answer questions such as what sequence of behaviors have the highest hit rate? What sequence of behaviors is most often used?

## 2.1 Methods

*Participants*

A sample of 48 infant-caregiver dyads were initially recruited from middle class neighborhoods in San Diego County. The participants were recruited by word of mouth at a postpartum exercise class', and posts on parent listservs and playgroups. Infants were excluded from the participant pool if they had been born over 2 weeks premature or had significant prenatal complications, or sensory or neurological problems. In addition, 5 families withdrew, so the final pool of participants who participated consisted of 43 infants (20 female, 23 male) along with each of their biological mothers who were all fluent English speakers. These participants participated in data collection monthly, from 6-9 months of age.

When creating the dataset using any coded behaviors, some mother-infant dyads were dropped each month due to the lack of clean data. As a result, the final number of dyads available for further data analysis was 37 in 6 months, 29 in 7 months, 38 in 8 months, and 40 in 9 months.

Environment

Researchers visited participant's homes on a monthly basis, and collected data using three Canon Optura mini-DV camcorders. The home environment provided a more naturalistic setting and was chosen for that reason. Three tripods were set up in the room each holding a camcorder, covered in a beige cloth, and with an animal puppet attached as the target. The tripods were placed at different parts of the room to provide a front, side and back view of the infant. The animal puppets were controlled for familiarity and size of the puppets, age appropriation, and were considered complex and interesting for infants. The combination of puppets varied month to month, yet the combinations were identical for each participant in any given month. Two additional toys were also provided as distractors, providing a control for whether the infants were attracted to the toys, or were solely following their mother's cues instead

## 2.2 Procedure

Data collection

Once consent was obtained, the mother was asked to sit facing their infant in a specific position, to maintain consistency across sessions and participants. The mother was then instructed to engage in a free play session for six minutes with the two distractor toys. The toys were then removed, and the mother would begin attempting to initiate shared attention with the infant by directing their attention to the target puppets. The mothers were told to direct their infants' attention as they normally would. The researchers remained out of the room during the session unless the infant became fussy.

Coding

Videos were captured using VirtualDub software at 30fps, and videos were clipped at a common sync point. They were then downsampled to 10 fps for behavioral coding. Mangold INTERACT was used to code three minutes from the attention following sessions, and coders annotated for mother's manual actions and gaze shifts as well as the infant's actions and the location of infants' visual fixations. In addition, verbal content type coding was conducted on Excel.

Some relevant variables coded for in this study include verbal: attention, object name; and non-verbal: gaze and point. Additionally, the following were noted: the onset and offset time of each action, the target at which the mother was attempting to direct the infant's attention towards, and whether the mother was successful in directing the infant's attention towards the target (hit). More detailed definitions of relevant terms are below in *Table 1.1*.

| **Term** | **Definition** | **Example** |
|---|---|---|
| Target | The puppet mother is attempting to direct the infant's attention to (frontcue, midcue, backcue) | ** tiger puppet -> midcue |
| Gaze (L) | If mother looks at target puppet | ** mother looks at tiger puppet |
| Point (P) | If mother points at target puppet | ** mother points at tiger puppet |
| Attention (a) | If mother uses attention directing language | "Look over here!" |
| Object Name (n) | If the mother names target puppet | "tiger" |
| Onset | Time stamp in video when mother starts bid action | Start frame: 990 |
| Offset | Time stamp in video when mother ends bid action | End frame: 1005 |
| Hit | True if infant looked at correct target within (5)s of cue end and before any wrong targets | ** infant looks at target puppet |
| Id | PID of the mother-infant dyad, 42 total | Id 1 -> mother-infant dyad 1 |
| Age | Age of infant in months (6, 7, 8, 9) | Age 6 -> infant is 6 months old |

*Table 1.1: Provides definitions and examples for relevant terminology. Information for each term was collected from the videos in the study.*

Statistical Analysis

Statistical analyses were conducted using R (v.4.4.1). The purpose of the analyses was to determine whether a) certain cue counts (number of cues within bids) were more effective in

eliciting Joint Attention in infants (hit rate); and b) whether some cue sequences were more effective in eliciting Joint Attention in infants. Non-verbal cues consisted of: *Gaze/Look* - mothers turning their heads and fixing their gaze upon the target; and *Point* - mothers pointing at the target, and verbal cues consisted of: *Attention* - attention directing language used by the mother; and *Object name* - mother says the target's name. While there are many more verbal content types that were recorded in the study, *attention* and *object name* will only be included in analyses as they are most relevant to attention directing specifically.

A mixed logistic model was used to examine the relationship between cue count and hit rate. A Repeated Measures ANOVA was used to determine the significance of variability in hit rate between relevant cue sequences. Lastly, Post-Hoc Tukey Adjusted Comparisons were used to further explore the significance of variability in hit rate between all possible pairs of relevant cue sequences.

### 3.1 Results: Cue Count

#### *Verbal Cues Excluded*

The number of cues per bid excluding any verbal cues were first analyzed *(Figure 1.1)*. The number of cues per bid ranged from 1 to 13, with a notable decrease in the number of bids with 4 or more cues. To increase the power of the analyses, bids with 9 or more cues were dropped and bids with 4 to 8 cues were grouped together. Looking at the frequency of bids with 1, 2, 3, and 4-8 cues, bids with 2 cues were the most prevalent. No significant variability was observed longitudinally.

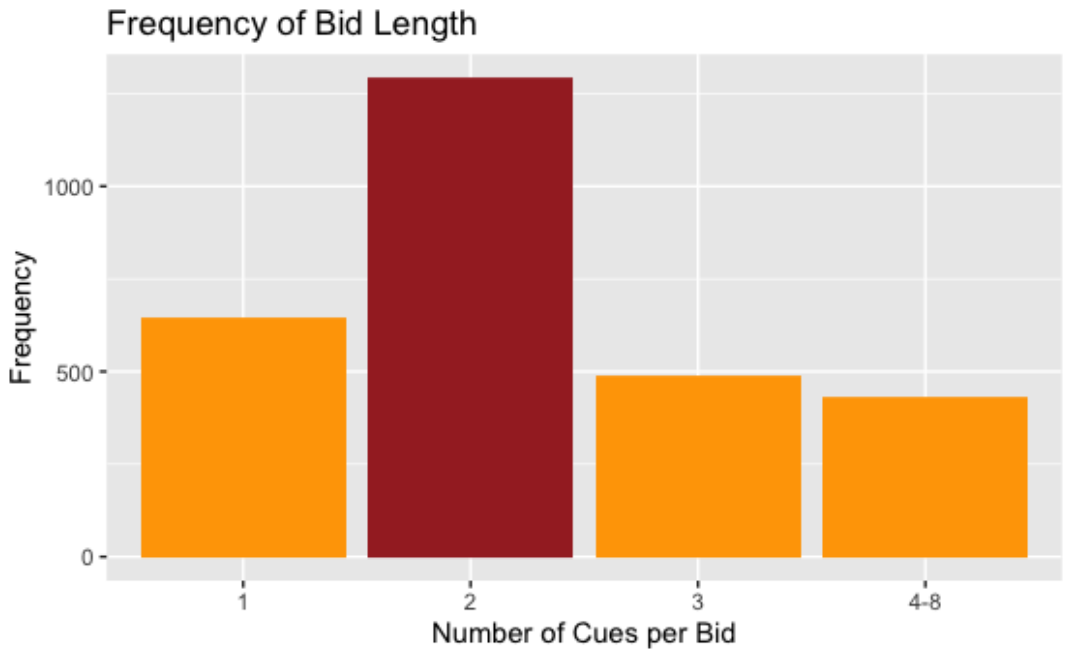


***Figure 1.1****: Frequency of cue count per bid, excluding verbal cues. Bids with 4-8 cues were grouped and bids with 9+ cues were excluded. Plot shows if mothers prefer bids with a certain number of cues over others. Mothers found to prefer bids with 2 cues over others. No significant longitudinal variation.*

The average hit rate of bids (1, 2, 3, and 4-8 cues) were then analyzed *(Figure 1.2)*, which revealed a general positive trend in hit rate as the number of cues increased. A mixed logistic model then analyzed the correlation between hit rate and cues per bid, accounting for variability in baseline hits across individuals. "Count" was used as the predictor and "hit" was used as the outcome variable. The model found a statistically significant ($p < .001$) positive correlation where a one unit increase in count is associated with a log odds of a hit increase by 0.395, which is equivalent to a 48% increase in odds of a hit for each additional cue.

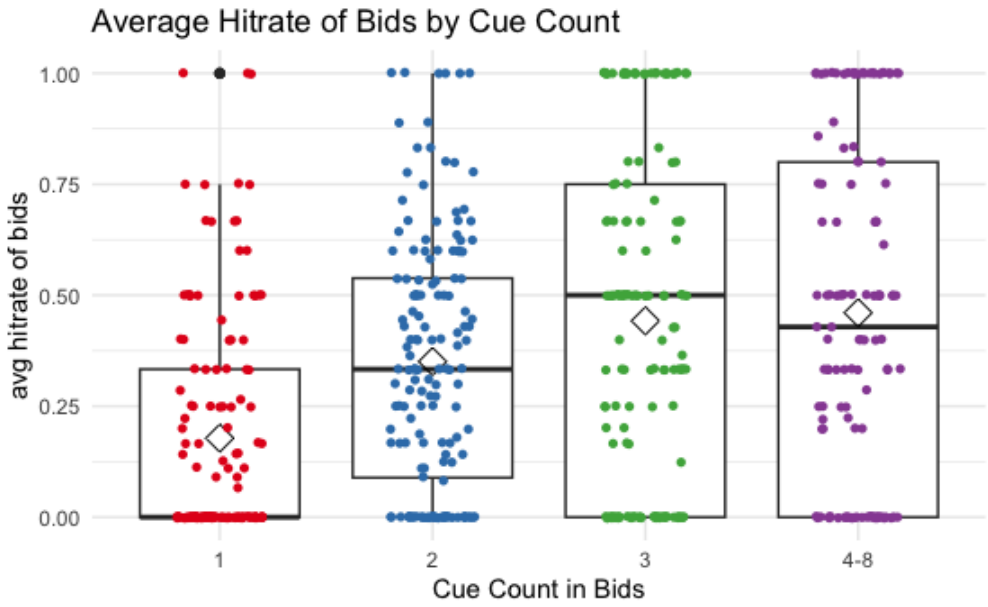


***Figure 1.2:*** *Hit rate of cue count per bid, excluding verbal cues. Bids with 4-8 cues were grouped together and bids with 9+ cues were excluded. Plot showed if there is any general trend between hit rate and cue count. The number of cues (count) significantly predicted the likelihood of a hit ($\beta = 0.395$, $SE = 0.046$, $z = 8.65$, $p < .001$), such that each additional cue was associated with higher odds of a hit. This corresponds to an odds ratio of approximately $\exp(0.395) \approx 1.48$. No significant longitudinal variation.*

## *Verbal Cues Included*

Next, cue counts per bid were examined with a dataset including verbal cues *(Figure 2.1)*. The number of cues per bid ranged from 1 - 14, yet similar to the no-verbal analyses, bids with 5-11 cues were grouped together and bids with 12 or more cues were excluded to increase power. The generated plot revealed that bids with 1 or 2 cues were the most prevalent, as opposed to the 2 cues per bid which was most commonly observed in the no-verbal analyses. In addition, the notable increase in frequency of bids with 1 cue suggests that the majority of 1 cue bids are verbal: *attention* or *object name*.

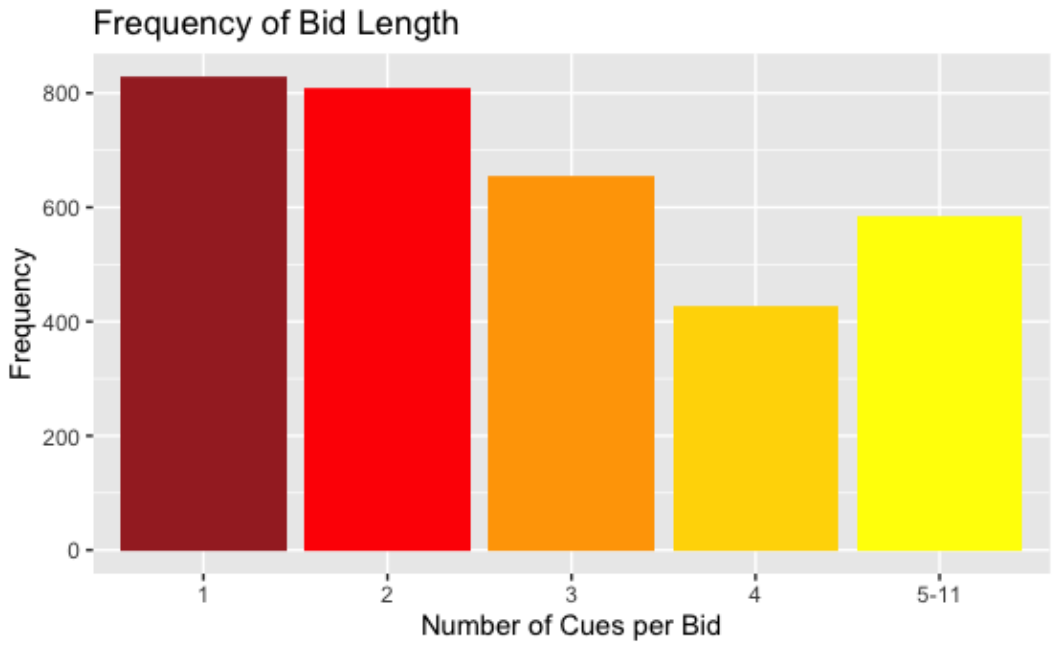


***Figure 2.1****: Frequency of cue count per bid, including verbal cues. Bids with 5-11 cues were grouped together and bids with 12+ cues were excluded. Plot showed if mothers prefer bids with a certain number of cues over others. Mothers found to prefer bids with 1 or 2 cues over others. No significant longitudinal variation.*

Lastly, the hit rate of cue counts per bid, including verbal cues, were examined *(Figure 2.2)*. Analysis revealed a statistically significant positive correlation between cue count per bid and the hit rate. A mixed logistic model was used, accounting for variability in baseline hits across individuals. "Count" was used as the predictor and "hit" was used as the outcome variable. The model found a statistically significant ($p < .001$) positive correlation where a one unit increase in count is associated with a log odds of a hit increase by 0.451, which is equivalent to a 57% increase in odds of a hit for each additional cue. Compared to findings where verbal cues were excluded, the mixed logistic model suggests a steeper positive correlation between cue count and hit rate. In addition, the smaller SE suggests a more precise estimate of the model's fit.

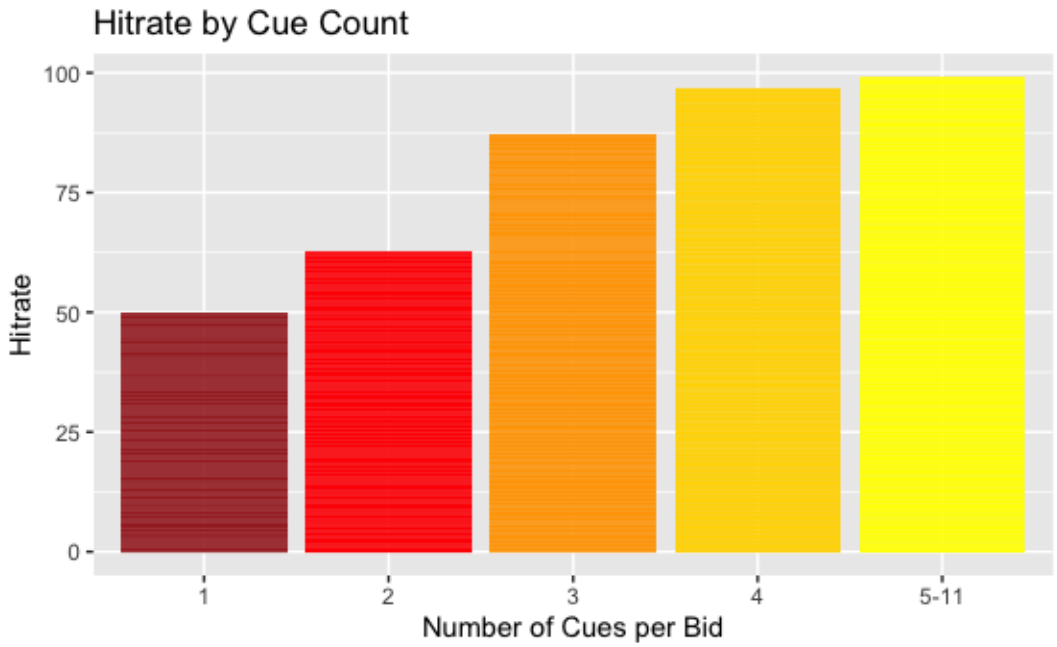


***Figure 2.2****: Hit rate of cue count per bid, including verbal cues. Bids with 5-11 cues were grouped together and bids with 12+ cues were excluded. Plot showed if there is any general trend between hit rate and cue count. The number of cues (count) significantly predicted the likelihood of a hit (β = 0.451, SE = 0.026, z = 17.40, p < .001), such that each additional cue was associated with higher odds of a hit. This corresponds to an odds ratio of approximately exp(0.451) ≈ 1.57. No significant longitudinal variation.*

### 3.2 Results: Cue Sequence

Cue sequences were next analyzed. While preliminary cue sequence analysis included non-verbal only analyses, only analyses which included both verbal and non-verbal data will be discussed, as they are more relevant to the purpose of this study. Cue types will be labeled as follows: L = Gaze/Look, P = Point, a = attention, and n = object name. A more in depth explanation can be found in *Table 1.1*.

First, the overall 5 most commonly observed cue sequences were extracted *(Figure 3.1)*. The "LP" was most prevalent out of the 5, and the only cue sequences containing verbal cues were the attention-only bid and the object name-only bid. The 5 most common cue sequences were then examined longitudinally (6, 7, 8, and 9 months). While the 5 most common cue sequences remained identical between the 7 month, 8 month, 9 month and the overall case, the 6 month distribution revealed that the "LPL" cue sequence was replaced with an "aLP" sequence.

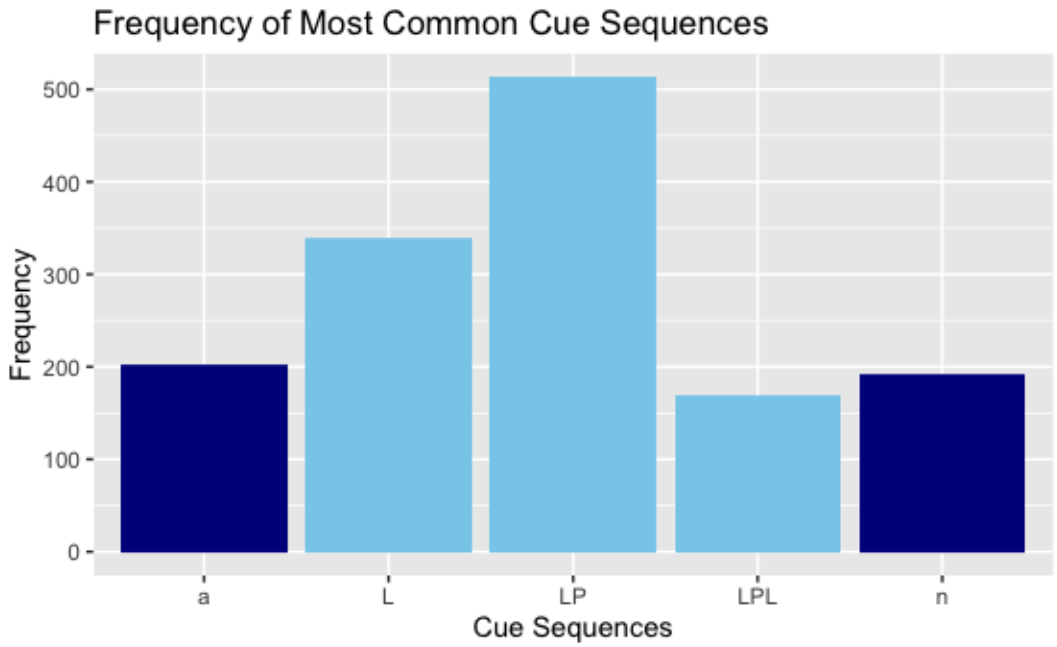


***Figure 3.1:*** *Frequency of identified cue sequences. L = look/gaze, P = point, a = attention, n = object name. Plot shows if mothers prefer a certain sequence of cues over others. Mothers found to prefer the "LPL" cue sequence over other sequences. Cue sequence "LPL" replaced by "aLP" in 6 month distribution.*

The next step consisted of extracting hit rates for each of the 5 most common cue sequences. For hit rate analysis, any verbal-only bids such as attention-only "a" or object name-only "n" were excluded since "hit" were not assigned to those bids in the dataset. As a result, cue sequence - hit rate analysis will only be looking at the 5 most common cue sequences, excluding verbal-only cues.

The new distribution of the 5 most common bids consisted of: "L", "LP", "LPL", "LPn" and "aLP". *Figure 3.2* displays the hitrates for the 5 most common bids without controlling for age. In this case, it can be observed that "LPn" and "aLP" sequences appear to have the highest hit rate. A repeated measures ANOVA showed a significant main effect of cue sequence on hit probability, $F(2.63, 57.83) = 14.63$, $p < .001$, Greenhouse-Geisser corrected ($\varepsilon = 0.66$), indicating that hit probability varied significantly across cue sequences. Mauchly's test indicated that the assumption of sphericity was violated for cue sequence, $W = 0.233$, $p = .0005$, so Greenhouse-Geisser correction was applied. Post-hoc Tukey-adjusted comparisons revealed that "aLP" and "LPn" sequences resulted in significantly higher hit probabilities compared to "L", "LP", and "LPL". Specifically, "aLP" performed significantly better than "L" ($p = .0006$), "LP" ($p = .0018$), and "LPL" ($p = .0356$), but not different from "LPn" ($p = .999$). The lowest performance was observed for "L", which was significantly worse than "LPL" and "LPn".

The hit rate of the 5 most common cue sequences was also examined longitudinally. While cue sequences were identical between the over-all, 6 month and 9 month cases, and the distributions between the 3 were relatively similar, the 7 month and 8 month cases had some differences. Both the 7 month and 8 month distribution only included the "LPn" sequence, and the "aLP" sequence was replaced by "LPLL" for 7 months and "P" in 8 months.

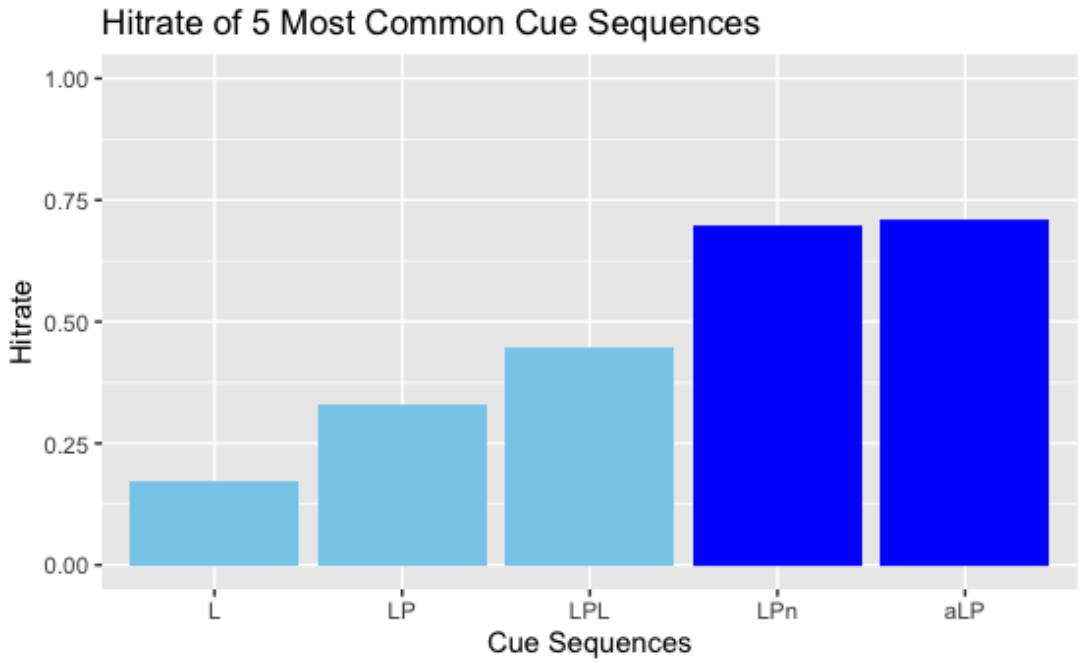


***Figure 3.2****: Hit rate of 5 most common cue sequences, excluding verbal-only cues. Plot shows if some sequences have higher hit rates. Repeated Measures ANOVA: significant main effect of cue sequence on hit probability, ($F(2.63, 57.83) = 14.63$, $p < .001$), Greenhouse-Geisser corrected ($\varepsilon = 0.66$). Mauchly's test indicated the assumption of sphericity was violated for cue sequence, ($W = 0.233$, $p = .0005$), Greenhouse-Geisser correction was applied. Post-hoc Tukey-adjusted comparisons: "aLP" and "LPn" sequences resulted in significantly higher hit probabilities compared to "L", "LP", "LPL". Lowest performance was observed for "L".*

The last predictor that was considered in this study are start cues, the first cue in the sequence of cues within each bid. The 4 possibilities include verbal starts: "a" or "n" and non-verbal starts: "L" or "P". The overall distribution combining all ages revealed a predominance of look/gaze-start bids *(Figure 4.1)*. No significant difference in distributions found longitudinally.

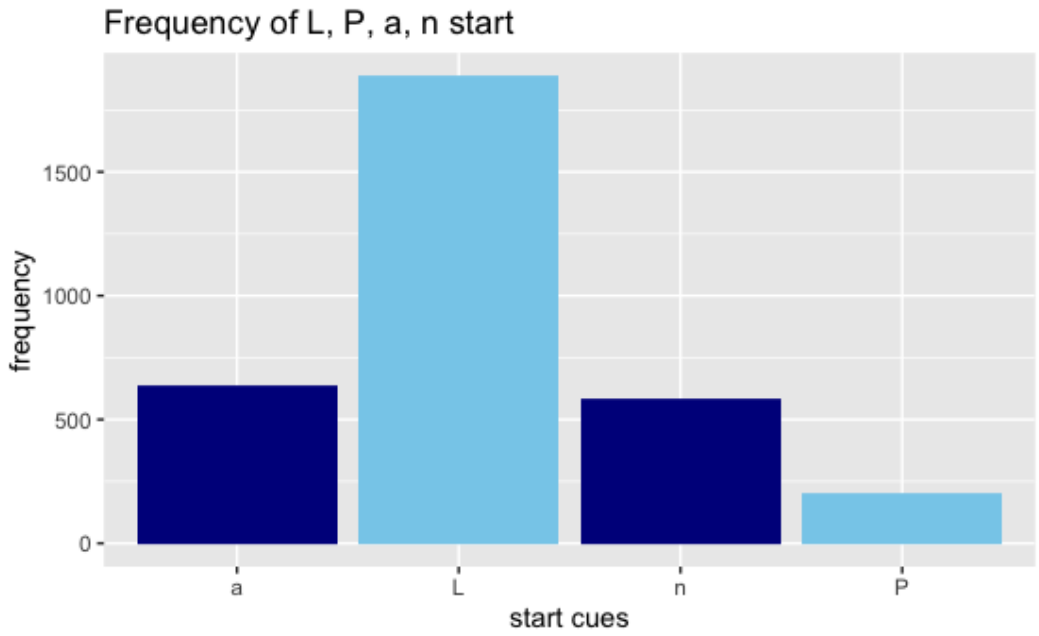


***Figure 4.1****: Frequency of different start cues. L = look/gaze, P = point, a = attention, n = object name. Plot shows if mothers prefer to start bids with a specific cue type. Mothers found to predominantly start bids with a look/gaze "L". No significant longitudinal variability.*

Finally, hit rates of the different start cues were analyzed *(Figure 4.2)*. Through conducting a Repeated Measures ANOVA, Mauchly's test indicated that the assumption of sphericity was not violated for age, W = 0.63, p = .89, or first cue, W = 0.11, p = .16. Yet for transparency when Greenhouse-Geisser and Huynh-Feldt corrections were applied, the effect of the first cue remained significant (GG-corrected: F(df1, df2) = value, p = .046). Post-hoc Tukey-adjusted comparisons revealed that sequences beginning with "a" were significantly more likely to result in hits compared to those beginning with "L" (p = .0062) and "P" (p = .0081), indicating that there was a significant effect of the initial letter of the cue sequence on hit probability. No other pairwise comparisons reached statistical significance.

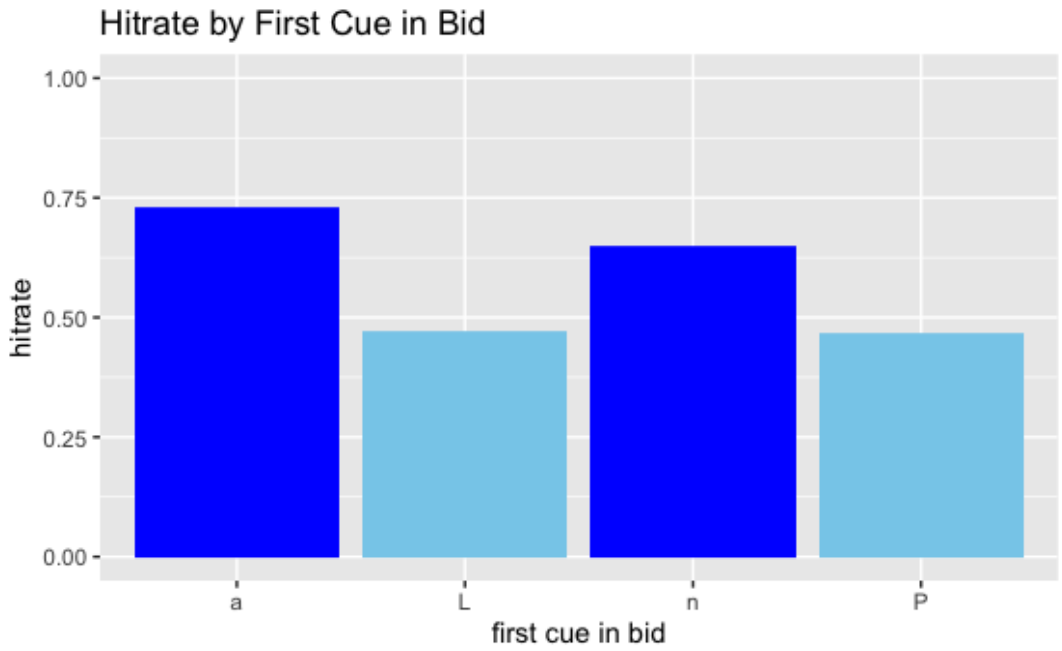


***Figure 4.2****: Hit rate of different start cues. Plot shows if some start cues have higher hit rates. Repeated Measures ANOVA: (GG: p = .046, HF: p = .032, p < 0.05), Post-Hoc Tukey-Adjusted Comparisons: sequences beginning with "a" significantly more likely to result in hits compared to those beginning with "L" (p = .0062) and "P" (p = .0081). No other pairwise comparisons reached statistical significance.*

### 4.1 Discussion

While there has been extensive research done on Joint Attention in general, there have been limited studies that investigate maternal cue sequences and their impact on eliciting Joint Attention. This is an important avenue to consider as it may give researchers a better idea of specifically what cue sequences are most impactful. Flom and Pick (2004) showed that the addition of verbal cues don't seem to significantly impact the frequency of hits or “successful” attention following, but did prolong joint attention. Yet it poses the question of whether the addition of verbal cues in specific cue sequences can result in a significant increase in the probability of engaging in Joint Attention with an infant. Or is the number of cues the mother uses to direct their infant’s attention more relevant compared to the type of cue used and their sequence.

This study investigated maternal cues used to elicit Joint Attention to determine whether some cue sequences are more successful in eliciting Joint Attention than others. More specifically, the number of cues used in cue sequences, and the unique sequences used were analysed.

## *Cue Count*

There are varying factors to consider when looking at bids, yet one prominent variable that hasn't been sufficiently investigated is the number of cues within bids. While some bids may only contain a single maternal action, others can contain 8, 9, 10 and so on. So why do some bids contain more cues than others? Do mothers tend to use a certain number of cues more commonly than others? Does increasing the number of cues positively impact the hit rate? Results from this study suggested that the number of cues within bids were positively correlated with the hit rate *(Figure 2.2)*. While findings were significant, preliminary analysis *(Figure 2.1)* showed a negative correlation between the number of cues within bids and their frequency. In sum, although mothers tend to use bids with less cues, bids with more cues appear to be associated with higher hit rates. One potential reason for longer cue sequences may be the mothers' felt need to reinforce successful Joint Attention with continuous cues. While the mothers were not instructed to maintain the infant's attention upon an object for an extended period of time, mothers may subconsciously feel motivated to keep the infant's attention upon the target through continuous cues. Another may be that the mother's continuous directive for the infant's attention to shift to a specific object was more effective in communicating the mother's goal. While the infant may have been distracted during the first few cues, or they were unsure what the mother was trying to accomplish, a longer sequence of cues may give the infant more time to process what the mother is asking for. These findings may serve as a motivator for further investigation upon the role of cue counts on eliciting Joint Attention.

## *Cue Sequence*

The majority of bids were found to contain 1 or 2 cues, yet what exact cue(s) are these bids made up of? Could some specific sequence of cues be more effective in eliciting joint attention than others? *Figure 3.1* suggests that the "LP" sequence was most commonly used by maternal figures, yet *figure 3.2* shows that "aLP" and "LPn" sequences were most successful in eliciting joint attention. More specifically, "aLP" performed significantly better than "LP" ($p = .0018$), regardless of the prevalence of the "LP" sequence. These findings may suggest that the addition of verbal cues, specifically attention-language and object name onto the non-verbal sequence "LP" can increase the hit rate. This could be an interesting finding to further investigate as it may provide a better understanding as to whether the addition of verbal cues is only successful in prolonging Joint Attention, whether verbal cues are effective as support for some cue sequences, and whether verbal cues are redundant in others.

This study also found that infants may be more likely to engage in shared attention upon an object when the first cue used is an attention-directing verbalization. Studies have found that in a more naturalistic setting, using verbalizations to elicit attention can lead to greater frequency in infants looking towards their caregivers, compared to gaze alone (Deák et al. 2008). Considering

such previous findings, starting a cue sequence with an attention-directing verbalization may serve as both an attention getting and directing cue, potentially leading to an increase in successful episodes of Joint Attention.

### *Limitations and Future Direction*

While the positive relationship between cue count and hit rate was found to be significant, future studies could benefit from a larger sample size. In this study, bids with 1 - 4 cues, when considering both verbal and non verbal cues, were most commonly observed, which resulted in the need to combine bids with 5 - 11 cues and entirely reject bids with over 11 cues from the study. Increasing the sample size may provide opportunities to independently consider bids with 5 or more cues, as well as a larger variety of cue sequences without the need to consider the study's power.

As mentioned above, *Figure 3.2* indicates that the "aLP" and "LPn" cue sequences fall in the top 5 most commonly observed cue sequences, when excluding verbal-only cues, and that the 2 cue sequences had significantly higher hit rates compared to "L", "LP", "LPL". In this case, it appears that the addition of verbal cues upon the cue sequence "LP" corresponded with an increase in hit rate. While no correlations can be concluded from this singular case, this speculation may provide reason to further examine whether the addition of verbal cues upon specific cue sequences is correlated with an increased hit rate. Similarly, while no findings from this study have suggested such correlation, it may also be interesting to control for the number of cues within bids, and examine whether specific cue sequences of the same cue count result in a higher hit rate.

Start-cue analysis revealed that attention-start cue sequences were significantly more effective in eliciting Joint Attention. Yet, it cannot be concluded that attention-start cue sequences lead to higher hit rates, as there is the potential that the hit rate is due to the existence of the verbal cue within the sequence and not necessarily its position in the sequence. A larger dataset will be required for such analysis, as the current data set did not provide a large enough sample of sequences containing both verbal and non-verbal cues.

Lastly, it would be beneficial to re-run all analyses using a complete data set which contains "hit" information for verbal cues to strengthen findings. The current dataset did not contain this information, and as a result "hit" was assigned through timing information between verbal cues and non-verbal cues, which were assigned True or False for "hit".

While this study was able to dig deeper into the potential implications of cue sequences on successfully eliciting Joint Attention, there is still much room for further research. Findings from this study suggest a potentially positive correlation between cue count and hit rate, yet with the

need to confirm this correlation with a larger data set. Findings also suggested that the addition of verbal cues upon otherwise non-verbal cue sequences may be associated with a higher chance of engaging in Joint Attention. This suggestion may provide a foundation for further investigation which will need a larger sample size as well as be controlled for cue length and non-verbal cue sequences. Lastly, results indicated that attention-start cue sequences were more effective in eliciting joint attention as opposed to look-start or point-start, yet this finding also requires further investigation as to whether the position of the verbal attention cue in the cue sequence is relevant to determining the effectiveness of the bid.

**Acknowledgements**

I would like to thank my PI Dr. Gedeon Deák for suggesting I pursue an Honors Thesis, providing me with the tools and resources I needed to succeed, and the opportunity to learn so much about the field of child cognitive development and the research field in general. I would like to thank Dr. Shannon Ellis for their insight on data analysis, and Yueyan Tang for meeting with me countless times to work out bugs while coding and helping me stay on track with my project. I would like to thank the UCSD Cognitive Science Honors Program Chair: Dr. Bradley Voytek for their mentorship throughout the program and their incredible presentation tips and tricks. Lastly, I would like to thank the rest of the Cognitive Development Lab who made this project possible, the families in SD who participated in this study, and my family and friends for supporting me throughout this journey.

**Sources**

Baldwin, D. A. (1991). Infants' Contribution to the Achievement of Joint Reference. Child Development, 62(5), 875 - 890.

Bard, K. A., Keller, H., Ross K. M., et al. (2021). Joint Attention in Human and Chimpanzee Infants in Varied Socio-Ecological Contexts. *Monographs of the Society for Research in Child Development. 86*(4).

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*, 10(3), 169–175.

Deák, G. O., Krasno, A. M., Jasso, H., & Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. *Infancy*, *23*(1), 4-28.

Deak, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental science*, *17*(2), 270-281.

Deák, G. O., Walden, T. A., Kaiser, M. Y., & Lewis, A. (2008). Driven from distraction: How infants respond to parents' attempts to elicit and re-direct their attention. *Infant Behavior and Development*, *31*(1), 34-50.

Flom, R., Deák, G. O., Phill, C. G., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development*, *27*(2), 181-194.

Kaplan, B. E., Yu, C. (2024). Multimodal Pathways to Joint Attention in Naturalistic Contexts. *IEEE International Conference on Development and Learning (ICDL).*

Striano, T., Chen, X., Cleveland, A., & Bradshaw, S. (2006). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology*, *3*(3), 289–299.

Tang, Y., Deák, G. O. (2024). Multimodal Maternal Input: Exploring the Dynamics of Joint Attention in Naturalistic Infant-Caregiver Interactions. *IEEE International Conference on Development and Learning (ICDL)*

Tang, Y., Gonzalez, M. R., Deak, G. O. (2023). The slow emergence of gaze- and point-following: A longitudinal study of infants from 4 to 12 months. *Developmental science*, *27*(3).

Tomasello, M., & Farrar, M. J. (1986). Joint Attention and Early Language. *Child Development*, *57*(6), 1454–1463.