Investigating the impact of different nonverbal and verbal cues on joint attention interactions between infants and caregivers

Angela Liu

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

University of California, San Diego

Cognitive Science Honors Program, 2023 - 2024

Background

Within their first year, infants develop new social and cognitive skills and interact more – and in more complex ways – with their caregivers as well as with inanimate features of their physical environments. An important social-cognitive skill that involves these interactions is being able to engage in joint attention. *Joint attention* occurs when two or more individuals converge their attention onto the same target, as a result of social cues being used to direct attention. This skill starts to emerge around 6-12 months of age, and initially appears in the form of *attention following* (AF) in young infants (Deák et al., 2000). AF can be defined as the searching and locating of targets based on cues from another individual's point or gaze (Tang et al., 2023). As infants develop, they use AF to engage in more *triadic interactions*. Triadic interactions occur as infants are engaging in attention with another individual and an object of mutual interest (Brandone et al., 2019). These triadic interactions demonstrate that infants have developed beyond the ability to only focus on face to face interactions, but can now also focus on other objects. Being able to engage in triadic interactions requires infants to effectively shift attention between social partners (e.g., caregivers) and objects (e.g., toys, or tools such as bottles or spoons) (Bakeman & Adamson, 1984, Newson & Newson, 1975). This ability to alternate attention between a person and an object allows infants to engage with and share new information with others (Bakeman & Adamson, 1984). Joint attention has also been shown to serve as a predictor or even a facilitator of other important social skills. For example, the prevalence of joint attention can predict trajectories of language development (Dawson et al., 2004), of learning associations between object labels and objects, which can help with object categorization, (e.g., Striano et al., 2006), and aid in the understanding of their caregivers' attitudes or reactions (e.g., Feinman, 1982).

Many studies have looked at the effects of different types of behavioral cues of caregivers that can precede infants' AF. Most studies have examined the effect of *pointing gestures* and *gaze shifting* when redirecting infants' attention. Gaze shifting occurs when someone shifts their line of sight from one point to another, often drawing attention to a target. With respect to gaze cues, previous research suggests that between ages 6-18 months infants increasingly respond to and redirect their attention when a caregiver silently shifts their gaze away from them to a distal target (Butterworth & Cochran, 1980). However, the infant's accuracy in fixating the intended

target seems to depend on variables including what is in the infant's visual field, the location of the target relative to the infant, properties of the target, and the magnitude of the gaze shift (Deák et al. 2000), and the caregiver's other behaviors (Deák et al., 2008). It's also important to note that although caregivers may use gaze shifts purposefully to communicate information, most gaze shifts are not primarily communicative, and infants do not usually respond to gaze-shifts alone (Deák et al. 2008, 2018). It has also been suggested there would be a difference in resulting AF when caregivers utilize pointing. Pointing gestures can be defined as "extending an arm and hand or finger(s) toward a target, to draw another person's attention to that target" (Deák et al. 2018, Kita 2003). Research has shown that infants, around 9-12 months old, can follow pointing gestures (Deák et al., 2008; Flom et al., 2003) Studies have suggested that pointing may elicit attention due to it being a more salient motion and being more intentional of an action, which may also be more helpful in a more cluttered environment (Deák et al., 2008; Deák et al., 2018). However, infants do not constantly monitor their caregiver's head position or face (e.g., Deák et al., 2014; Yu & Smith 2013), so multimodal engagement may be more likely to elicit AF, as this type of engagement draws on multiple modes of cues to draw attention (e.g. toy play that involves auditory cues and visual cues).

In Flom et al. (2004), the effects of conditions of gaze alone, gaze and point, and gaze, point and verbalizations were investigated. They found that using pointing in addition to gaze facilitated AF, in 9 month old infants. This might be because pointing and gaze act as redundant cues, supporting each other as spatial cues specifying the same target region. However, there was less of an effect from adding verbalizations to point and gaze. This finding about the effect of verbalizations in AF seems to support Flom and Pick's (2003) finding that additional verbalizations didn't increase the frequency of 9-month-olds' AF. However, their results suggest that verbal cues can prolong joint attention, which raises a question about the role of verbalizations in joint attention episodes. That is, are point and gaze cues more useful in re-directing an infant's attention, whereas verbal cues are effective in sustaining attention?

The role of verbalizations has been further examined in more naturalistic settings. Most research studying point and gaze, or more generally joint attention, has been done in laboratories where there are fewer distractions in the environment around the infant. However, at home, it's

likely that having a less-controlled environment may present more distractions, with clutter and other objects that can attract an infant's attention. When redirecting their infant's attention, caregivers may first attempt to elicit their attention, then attempt to re-direct it towards a specific target. 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). While an infant shifting their attention to their caregiver doesn't necessarily lead to AF, it can be an important initial step that facilitates directing attention. Not only is the presence of speech an important variable, but the content of the speech may also play a role. Caregivers use a variety of words that can be categorized into content types, when talking to an infant. The specific content of maternal speech may depend on context and where the infant's attention was directed towards. For example, research has suggested that when infants are handling toys, maternal speech may contain more object name labels and descriptive language (Chang & Deák, 2019). When infants shift their gaze away from faces or toys, infants may hear their own names and attention-related language (Chang & Deák, 2019).

In this research project, we will examine the impact of specific verbal *content types* used by mothers, in addition to point or gaze, when they attempt to direct infants' attention to a distal target. The content types we will examine are: infant's name, object (target) label, and attention-eliciting language. As previously mentioned, caregivers may use an infant's name to elicit attention when their gaze has shifted away from a face or object (Chang & Deák, 2019). I believe it's important to consider why their name can be a reliable cue, to at least elicit attention in AF. Many studies have shown that infants, from a very young age, can respond and recognize their name (Csibra, 2010) and may even prefer to hear their name over other names (Mandel et al., 1995). When considering how infants may respond to hearing object labels in AF tasks, other research suggests that shared attention helps facilitate object label learning. In a study with 18 month old infants, research has shown that when parents named a visually dominant object, their infants were able to learn the object label (Yu & Smith, 2012). This evidence prompts a question if infants may respond more reliably to AF when maternal speech contains an object label, as they may be building association between the object label and the referent. And while infants may be too young to understand the semantic meaning of object labels, do AF tasks help build representations of objects? The last content type considered is attention related language, which

contains words that may have the clearest expectation for attention, when used with older children and adults. For example, when someone says “look!”, there is an underlying expectation to look or direct attention towards a certain object or target. However, when personal names or object names are used (in isolation), they may carry different meanings, and may just be more context dependent. Again, while infants may not yet understand the semantic meaning behind attention-related language, can they build associations in the contexts in which they hear attention-related language and their caregivers’ expectations for their responses?

Whereas we predict a higher frequency of instances where mothers use verbalizations in their bids to redirect infants' attention, we also predict that infants will increasingly reliably respond to specific labels or verbal commands as they age from 6-9 months of age. It is possible that the development of these associations lead some younger infants to respond to specific words by looking for an appropriate referent. In addition to object labels and the infant's name, can infants learn what expectations caregivers may have when using attention related language like “look”?

Method

Participants

A sample of convenience of 48 infant-caregiver dyads were recruited from middle class neighborhoods in the San Diego county. They were recruited through word-of-mouth, postpartum exercise classes, and posts to parent listservs and playgroups. Caregivers used in the data collecting sessions were all the infants’ biological mothers, and were all fluent in English. Infants were excluded from the sample if they were >2 weeks premature or had significant prenatal complications or sensory or neurological problems. Five families withdrew from the study, and were therefore excluded from the data analyses. There were 43 remaining infants (20 female, 23 male), who participated in data collection monthly, from 6-9 months.

When creating the larger combined dataset, that included information for behavior during AF tasks, as well as speech data, the number of usable participants changed due to availability of clean data. When doing analysis, there were 37 participants in 6 months, 29 participants in 7

months, 38 participants in 8 months, and 40 participants in 9 months. More detail about the data cleaning and organization is provided in the Results section.

Materials

Figure 1. Home experimental set up. Mother and infant are sitting facing each other, and there are three animal puppets placed throughout the room, in front, side, and back of the infant. There are also distractor puppets placed on the top of the targets.

Test environment: Home

Researchers went to the participants' homes to collect data, with the home providing a more naturalistic setting. Infants were seated comfortably in commercially available walker, in the typical play area of their home. Mothers sat on a cushion on the floor, across from the infant, so they were facing each other. The walker's height was adjusted so the infant's and mother's eye levels were approximately equal. There were three Canon Optura mini-DV camcorders on tripods that captured three different perspectives of the session, showing the front, side, and back of the infant.

There were four tripods covered in beige cloth, placed at predetermined angles and distances from the infant and mother. These tripods were used to hold the camcorders and

distractor toys. There were also three animal puppets, functioning as targets, placed in front of three of the tripods.

Toys

As previously mentioned, different animal puppets were used as targets in each session. The familiarity and size of each puppet was controlled, were age appropriate, and considered complex and interesting for infants. Each target puppet was placed upright against the beige cloth covered tripod, as the cloth provided a neutral colored background. Each set of three target puppets was kept controlled for every participant in a given month. However the exact combination of target puppets differed from month to month.

Two additional toys were used as distractor toys, providing a control for whether infants were just generally attracted to toys, as opposed to following mothers' cues to specific puppets.

Procedure

Data Collection

After obtaining consent, infants and their mothers were seated in pre-specified positions and researchers left the room. After a six minute free play session (described in another study, de Barbaro et al., 2016), the AF session began as mothers attempted to direct their infant's attention to the target puppets. During the AF session, there were no toys and had been removed by researchers before. At the start of the session, mothers were told "try to get your child to look at all three of the puppets in front, to the side of, and behind you. Do this how you would normally direct your infant's attention to an object". Researchers subsequently remained out of the room for the entirety of the session, unless the infant became fussy.

Coding

Mothers typically started a bid by using a pointing gesture, this gesture was used to denote the beginning of the AF task. A bid is the term used to refer to when the mother initiates the AF task, attempting to direct the infant's attention towards any of the targets.

Videos were captured at 30 fps, and using VirtualDub software, videos were clipped at a common sync point, and downsampled to 10 fps for behavioral coding. Mangold INTERACT (Mangold, 2022) was used to code three minutes from the AF session, 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 (e.g. mother, tray table, targets).

As part of another coding project, maternal speech was also transcribed, synced with maternal manual actions and gaze shifting data. Coders then categorized the content in maternal speech into content types (based on Chang & Deák, 2019). Table 1 below shows the definitions of the content types, borrowed from Chang & Deák (2019), but for the purposes of this study, we only used attention language, infant name, and object name during AF tasks for the analysis.

Content type	Definition	Example	Example also contains
DECLARATIVE	Declarative statements, excluding one-word utterances	*That's pretty neat*	–
IMPERATIVE	Imperative syntax, including *Let's*	*Go get it*	ACTION
QUESTION	Question syntax, or utterance-final pitch rise in appropriate context	*What's over here?*	–
DESCRIPTION	Adjectives or predicates referring to toys	*It's the blue one*	DECLARATIVE
ATTENTION	Directing or commenting on infant's focus of attention	*Can you see mommy?*	QUESTION
ACTION	Directing or commenting on infant's non-perceptual actions	*You wanna try and squeeze it?*	QUESTION
OBJECT NAME	Contains name for physically present toys	*We got a little rubber ducky*	DECLARATIVE
INFANT'S NAME	Contains infant's name or other term of address	*Hi [name]*	SOCIAL ROUTINE
AFFIRMATION	Contains a form of *Yes* or acknowledges a conversational turn without further content	*Yeah*	–
SOCIAL ROUTINE	Fixed expressions such as greetings or exclamations	*Thank you*	–

Table 1. Definitions of each content type studied in Chang and Deák (2019).

Statistical analysis

All analyses were run using R [v.4.3.2].

Current analyses focus on understanding the combinations of nonverbal and verbal cues mother's use in these AF tasks to direct their infant's attention. In a given bid, mothers tend to use a variety of cues, as well as various combinations of these cues. Nonverbal cues include: (1) Pointing: mothers extending their arm and index finger to point towards a target; (2) Gaze: mothers turning their head and eyes, fixating on a specific target. Verbal cues included speech, any vocalization directed towards the infant, during the AF task. As mentioned in the background, analysis of verbal cues only considers the implications of specific content types in speech, infant name, object names (names of targets), and attention related language.

These analyses also include looking at the frequency of the combination of these cues, as well as the hit rates associated with these cues. The frequency rate was a proportion of bids that had a specific cue combination out of all bids initiated by a participant's mother, in a given month. A successful bid was defined based on "hit", which was if the infant looked at the correct target. The hit rate was a proportion of correct looks, out of all bids initiated by a mother.

Results

All analyses first aim to better understand maternal behavior and speech used in bids, towards infants to direct their attention to any of the targets.

While there were 43 participants with data from the AF tasks, the number of participants with usable utterance data and AF task data differed from month to month. For the purpose of the study, we retained the data from participants if they met certain criteria. First, participants have to have had recorded data from the AF task and usable speech data, meaning the speech was all transcribed, cleaned, and coded. However, due to the availability of video clips, there were some participants without speech data, so they were removed from the data set for subsequent analysis. After checking for this first criteria, there were 37 participants used in 6 months, 29 participants in 7 months, 38 participants in 8 months, and 40 participants in 9 months. While 6, 8, and 9 months had a similar amount of usable data, 7 months had fewer participants because the speech data had not yet been completely finished. However, despite 7 months having fewer participants than other months, we still see relatively similar trends in 7 months when compared to other

months. In our next steps, we would like to use all available data for 7 months for analysis, to be sure that having less participants didn't impact the results significantly. After obtaining the usable data, it was further cleaned to filter out bids where infants were already looking at the target. If the infant was already looking at the target, this would not be an example of a "hit" for the AF task, as the infant wasn't following cues to the target. This concluded the data cleaning process.

Different types of cues used in attention following

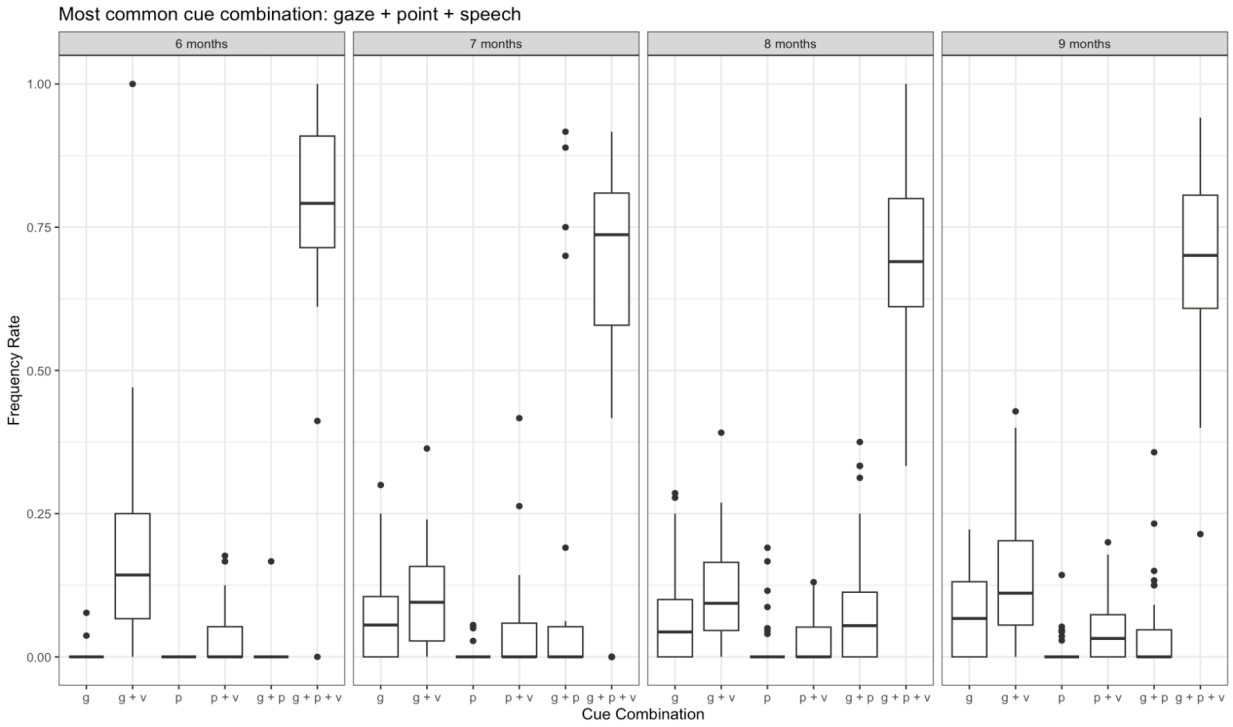

Figure 2. Frequency of each cue combination used in bids, in each month. There are six possible combinations, but the most common across all months is gaze, point, and speech.

First, we analyzed the frequency of each type of cues, combinations of nonverbal and verbal, that were used in each month. There were six total types of combinations that we looked at: gaze alone, point alone, gaze and point, gaze with speech, point with speech, and gaze and point and speech. At this point in the analysis, the verbal cues were treated as speech as a whole, not yet analyzing each content type.

The frequency in which these combinations of cues in a given bid are shown in the figure above. The frequency of each cue combination was calculated for each participant, in a given month. From these results, it's evident that the most common cue combination used by mothers is gaze, point, and speech, used together, across all months. In six months, the median frequency rate for gaze, point, and speech used together was about 79% of all bids, in seven months, about 74% of bids, in eight months about 69% of bids, and in nine months, about 70% of bids.

Specific content types in speech

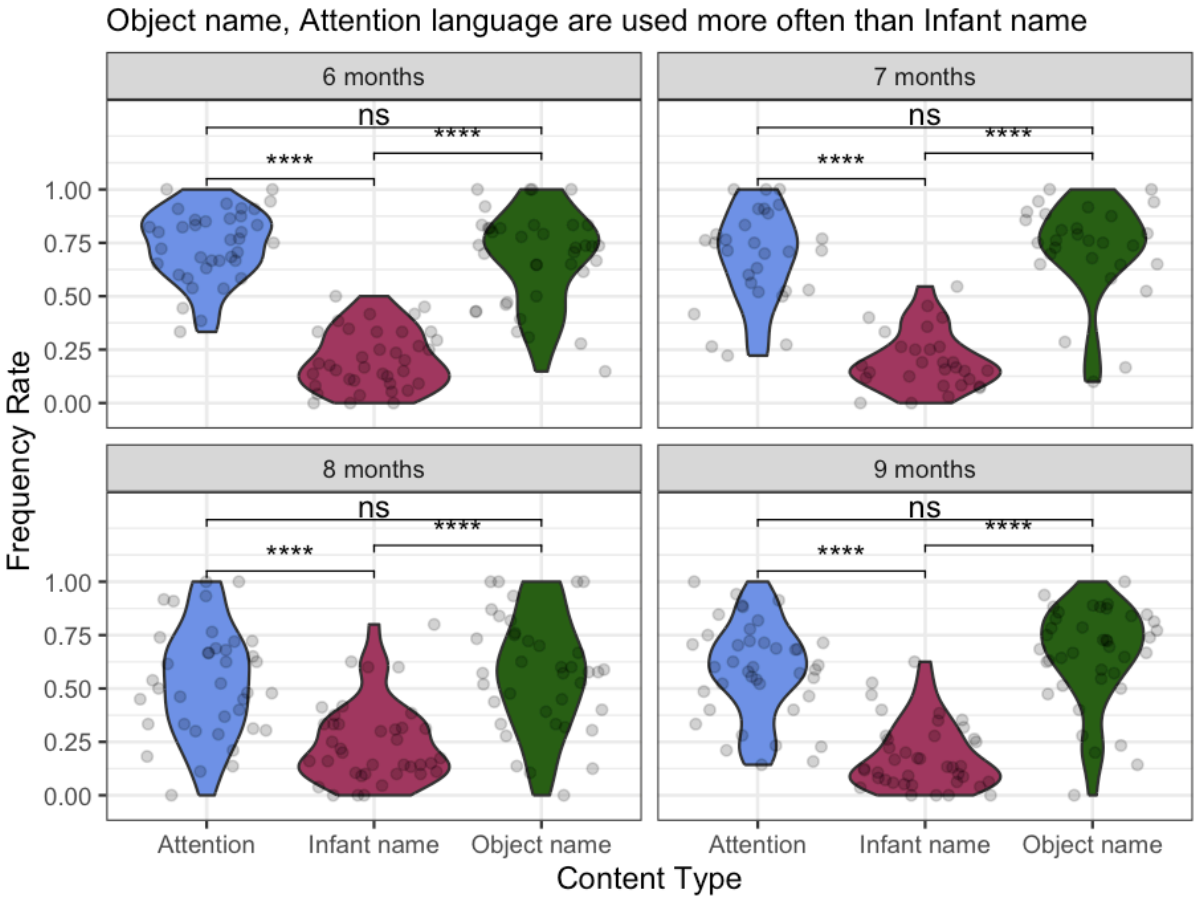

Figure 3. Distribution of each content type used in speech, for all months

Month	Attention language Median frequency	Infant name: Median frequency	Object name: Median frequency
6	0.76	0.18	0.73
7	0.71	0.15	0.76
8	0.48	0.17	0.58
9	0.58	0.13	0.69

Table 2. These are the median frequency rates for each content type for each month.

Month	Object name, Attention language p-value	Attention language, Infant name p-value	Object name, Infant name p-value
6	0.27	4.5e-13*	1.7e-11*
7	0.41	1.1e-09*	4.7e-09*
8	0.50	1.8e-06*	6.4e-07*
9	0.1	4.9e-11*	2.3e-11*

Table 3. These are the p-values when comparing the distributions of each content type, using the Wilcoxon Rank Sum test. Asterisks denote the p-values that suggest a significant difference in the distributions between content types.

Looking further at the content types, we used the data from speech transcriptions. Before any analysis, each bid had a combination of content type codes based on which content types present in a bid. The codes were based on definitions from Chang and Deák (2018). Because this study was specifically studying attention language, infant name, and object name, the analysis cannot account for statistics for other content types present, like questions, declaratives, and descriptive language.

To further investigate the specific relationship between attention language, infant name, object name, we first found the frequency in which these content types are used, including bids when they are used in combination with each other. For example, if the mother used both infant

name and attention language in a bid, that bid would count towards the frequency rate of infant name and attention language. The frequency rate was calculated to better control for differences in the number of bids used by participant's mothers. The frequency rate was the proportion of number of times bids had a specific content type out of total bids, for each participant. This was calculated for all months, and can be seen in Table 2. After doing a Wilcoxon rank sum test, there appears to be a very similar trend, where the frequency rate of object name and attention language aren't significantly different from each other, but there is a statistically significant difference when each is compared with infant name (specific p-values can be seen in Table 3 above). It appears that infant name occurs the least frequently in a bid, when comparing infant name, object name, and attention language. Another trend in the data is that the frequency of usage of attention language and object name seems to drop a bit when comparing the median frequency rate in 6 and 7 months, to 8 and 9 months.

Content types used in combinations, and hit rate

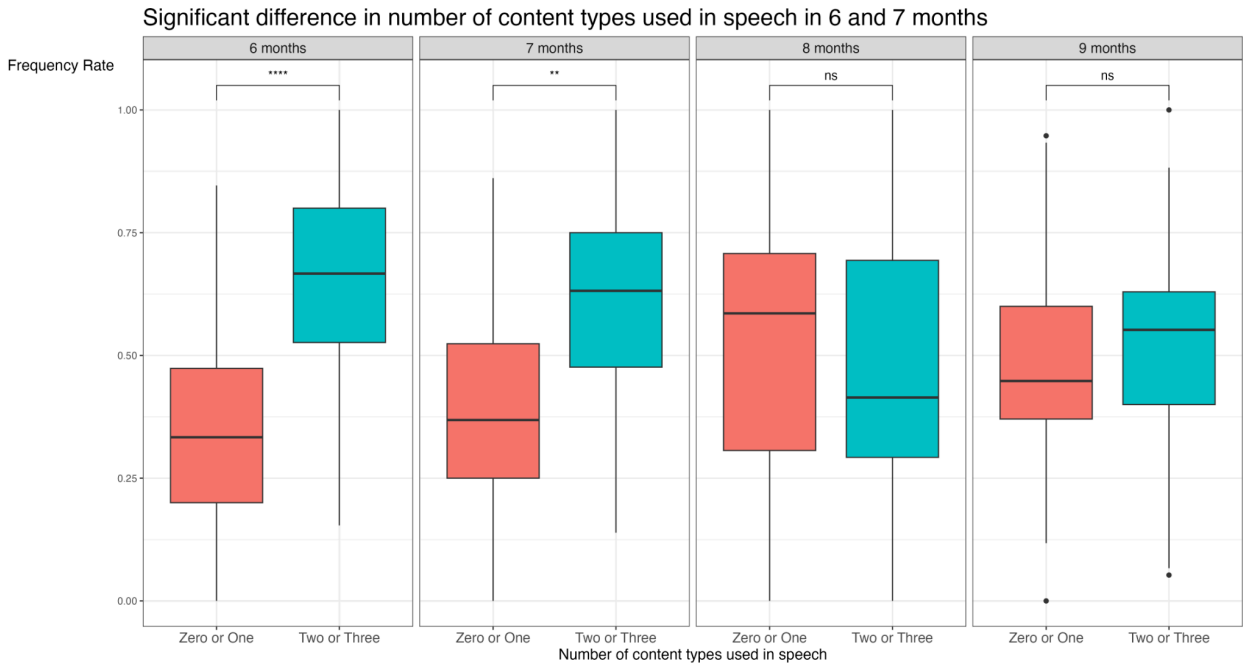

Figure 4. Frequency of the content type combinations in bids ("Zero or One" and "Two or Three"), in all months. Six and seven months show a significant difference in frequency of using zero or one content and the frequency of using two or three content types in combination.

To further understand the relationship between content types and AF tasks, we grouped the number of content type occurrences based on the combinations they may appear in. One group is the “Zero or One” group, meaning any of the content types investigated either weren’t used at all in that bid, or only one was used. An example would be, “Look at that!”, where the only content type, from the three being studied, is used in speech. Combining the zero and one group also provided more statistical power. The “Two or Three” group is when the speech from that bid had two or three of the content types used in combination. An example of this would be, “Angela, look at that monkey!”, where infant name, attention language, and object name were used. We analyzed this grouping for all months. The results show that after comparing the groups (using the Wilcoxon rank sum test), there is a significant difference between the number of content types used in combination in speech in six and seven months (with a p-value of 0.000013 and 0.0014 respectively), but not in eight and nine months. In both 6 and 7 months, mothers more frequently used two or three of the content types in combination when directing their infants in AF tasks. In 6 months, the median frequency rate of at least 2 content types being used in combination with each other was about 67% of the bids, whereas using at most 1 content type occurred around 33% of the bids. In 7 months, the median frequency rate of using at least 2 content types occurred in around 55% of bids, whereas using at most 1 content type occurred in around 46% of the bids. In 9 months, we also found that the frequency rate for the bids that were in the “Zero or One” category started to hover around the same rate as the “Two or Three category (starting to approach 50%).

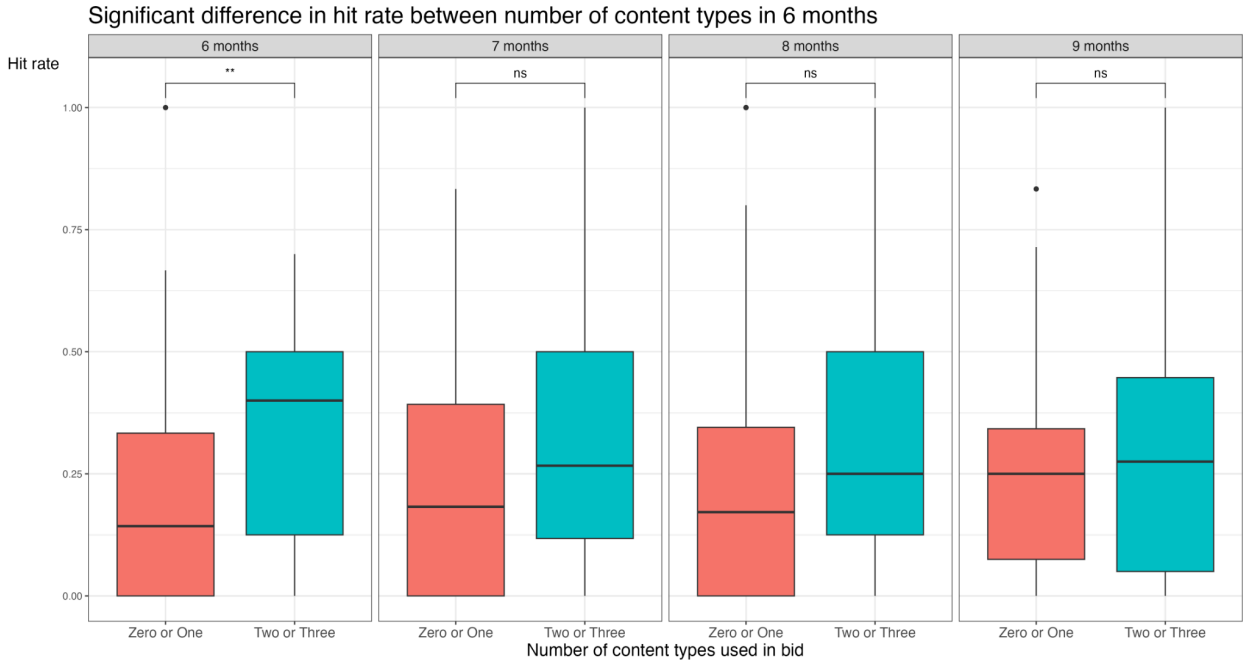

Figure 5. Hit rate associated with each content type combination. In six months, there is a significant difference in hit rate for when "Zero or One" content type is used in speech, compared to the hit rate for when "Two or Three" content types are used together.

As one of the research aims was to better understand how content types impact the "success" of AF, we also looked at how the number of content types used in combination was correlated with hit rate. Using the same groupings of "Zero or One" and "Two or Three", we looked at the hit rate associated with these combinations. The hit rate for each group was calculated by finding the proportion of hits from the number of bids that had zero or one content types or two or three content types, respectively. And this hit rate was calculated for each participant, for each month. After doing a Wilcoxon rank sum test to determine significant differences, the results show that there is only a significant difference between the hit rate between combinations of content types in six months (with a p-value of 0.0092), but not in seven through nine months. For six months, having at least two content types in the cue combination had a higher rate. The results also seem to show that the hit rate for "Zero or One" content type combination and "Two or Three" content type combination slowly becomes more similar as the infant gets older.

Overall hit rate

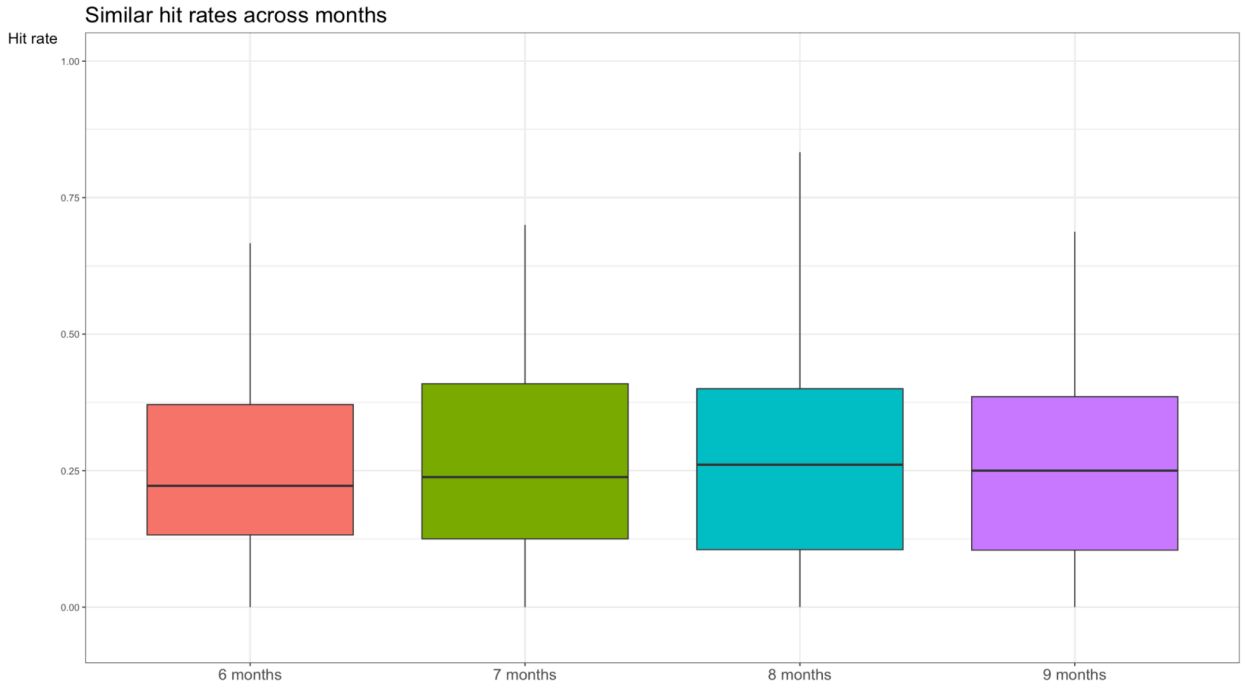

Figure 6. This is the overall hit rate for each month.

When disregarding the different cue combinations, the data shows the average hit rates across months is very similar. The average hit rate for each participant is the proportion of total hits out of all the bids for that participant. This hit rate was calculated for all participants, for all months.

Discussion

There has been much research on behavior preceding attention following and joint attention between infants and caregivers, especially on the cues that may be utilized to direct an infant's attention. There have been many studies that have looked at the point and gaze as cues to use in AF tasks (e.g. Flom & Pick 2003, Deák et al. 2008), finding that point and gaze together can act as redundant cues and better support attention following. The data on how speech impacts attention following and joint attention seems a bit unclear. Flom and Pick (2004) showed that adding verbalizations to cues didn't seem to significantly impact the frequency of

"successful" attention following, but did prolong joint attention. Our study's purpose sought out to better understand the relationship between specific nonverbal and verbal cue combinations and success in AF tasks.

Cue Combinations

When considering point and gaze, with verbal cues, our study also shows that mothers tend to use this combination of nonverbal and verbal cues to direct their infant's attention, more than other combinations. Without looking at how these cue combinations are associated with hit rate, this provides insight into what mothers think is helpful when engaging with infants. Seeing how much greater the frequency rate was for a cue combination of gaze and point and speech, it 's evident that mothers have preference for combination when directing attention. Even though our AF tasks are different from free play and we didn't study how long the infant's attention was held for, I think that it's possible that using both nonverbal and verbal cues together can support the idea that multimodal engagement helps create an interaction that attracts or holds an infant's attention. I think it's also possible that most of the time, even outside of AF tasks, using multimodal engagement is natural. Using both types of nonverbal cues, with speech, may be a choice that balances what is the most natural for the mother, as well as provides a mode for the mother to communicate information with her infant.

There has been some research on content types used in infant-mother triadic interactions (Chang and Deák, 2019), however the relationship between specific language used and attention is not yet fully understood. My research project sought to better understand this relationship, especially with the content types infant name, attention language, and object name.

Content types used in bids

Our results showed that in all months, object name and attention language were both used more often than infant name was. At first, this might seem surprising considering infant's learn from a young age the importance of their name (Csibra, 2010). If mothers have noticed that infants can respond to their name or use the infant's name so often to call them, why don't they use the infant's name in a bid to direct attention? I think the answer lies in the fact that names are often used to elicit attention, as opposed to directing attention. While we didn't analyze the cues

and speech content in the "before window" (before a bid), I would expect we would find a greater frequency of infant names being used. If the infant name was in speech in the "before" of an AF task, it was not counted in the frequency rate, as the results only showed the frequency rate of content types used in AF tasks. Since we defined a bid as starting with a point or gaze, I predict that mothers are calling the infant's name, to first attract attention, then use a point/gaze to start the bid. I think this also supports the idea that mothers are intentional with their word choice. They are choosing to use object names and attention language, when referring to the targets, which can help communicate more information. Using these content types during AF tasks may help create an environment and context where the infant starts to learn associations between different object labels and the objects themselves, as studies have shown that joint attention and AF supports language development (Dawson et al. 2004).

Our results also showed that when considering infant name, attention language, and object names, it's more common to use a combination of at least two or all of these content types in bids, in six and seven months. And considering object name and attention are used the most often in a bid, the results suggest that if mothers' aren't using some combination of all three, they may be using object name and attention language in the same bid. It's interesting to see how this significant difference in frequency only appears in six and seven months, but not in 8 and 9 months. When looking at Figure 4, it seems that the frequency rate of either two or three content types used in combination, or just one or none, is very similar in 8 and 9 months. I think the slight increase in frequency for "Zero or One" group (when looking at 6 and 9 months for example), doesn't mean that mothers are speaking less. In fact, we know that in eight and nine months, the most common cue combination is still gaze and point and speech, so mothers are still utilizing speech in bids in the later months. My hypothesis is that mothers are using the contents I looked at less, and are starting to expand and diversify the content in their speech as the infant gets older. The scope of this study is a bit limiting in that it only accounts for usage of infant name, attention language, and object names, but not other content types from Chang and Deák (2019). Mothers may be using more complex speech, and not just relying on object labels, names, and attention words.

Another important takeaway from this study is looking at the hit rates associated with the content types. In Figure 5, there wasn't a significant difference in hit rates, when comparing the "Zero or One" group and the "Two or Three" group. And when we looked at the overall hit rate across all months (Figure 6), there's a very similar hit rate across all months. At first, this was a bit surprising, considering there seems to be a preference for specific cues used, and preference for content types. I thought that if specific content types are used more often, infants can learn and react to these words sooner. However, looking at the overall hit rate, this may provide evidence that we are looking at a very young age, where infants are still developing the ability to attention follow. For example, in Figure 6, the median hit rates for each month are hovering around 25%, which isn't a very high hit rate. Even though mothers may be making a choice to communicate information to infants, these infants may be too young to yet have different responses to specific cues, as they are still learning to respond.

Limitations and Future Research

This project has some limitations, and given more time or for future research, I think it would be important to first use the completed 7 month speech data set. As mentioned earlier, at the time that we were analyzing the data, the 7 month speech data wasn't completely transcribed and cleaned, leaving only 29 usable participants. Even though the trends still held up, I think an important next step would be to re-run the analysis with the completed 7 months dataset.

Reflecting on this project, I think there are many factors that may also still be at play that impact an infant's ability to follow attention shifting. We limited the study to three content types, but as we hypothesize that mothers are diversifying their speech, it would be interesting to examine the trends of other content types. Do other content types have higher associated hit rates? Another question related to speech is how does the duration of speech impact joint attention. While this may be hard to compare from participant to participant, as mothers may inherently have different speaking rates, I wonder if having a longer sentence would lead to the infant being less interested, simply because the duration is too long. Or would having more words lead to more auditory stimuli, which could lead to longer interaction between the infant and caregiver.

Another important factor to further study is what is happening in this “before window”, before the bid. I think where the infant is attending to before a bid begins, as well as the cues the mother may utilize, may impact how much effort is required to follow attention. While we cleaned out the “already hits”, bids where the infant was already attending to the target, we didn’t analyze where the infant was attending before, if it wasn’t the target. It’s possible that where the infant is attending to before a bid may be more interesting, or farther away from where the mother is trying to direct attention to.

One other future direction I’m curious about is investigating how different content types are correlated with how long it takes an infant to shift their attention. It’s possible that individual differences can lead to different response rates, but I also wonder if specific content types lead to faster responses than others. While I predict that just based on the overall similar hit rate, there may be minimal differences in response times to specific content types, I think it would be interesting to also compare 6 months response rates to 12 or 18 months response rates.

From this project, we’ve been able to learn some more information about the ways infants may respond to different cues, as well as what decisions mothers and caregivers are making when engaging infants in joint attention interactions. However, we advocate for future studies on the specific relationship between content types and joint attention to better understand how to support successful attention following.

Acknowledgements

This project wouldn’t have been possible without all the guidance and support from my committee members, Dr. Gedeon Deák, Dr. Shannon Ellis, and Yueyan Tang. I’m so grateful for all of their patience and willingness to walk me through the research process, as a whole, as well as encouraging me to continue pursuing my questions. I would also like to thank the other members at the Cognitive Development Lab and the families in San Diego that participated in the study. Lastly, I want to express my gratitude to my family and friends who have supported me through all of my research endeavors.

Sources

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child development*, 1278-1289.

Brandone, A. C., Stout, W., & Moty, K. (2020). Triadic interactions support infants' emerging understanding of intentional actions. *Developmental science*, *23*(2), e12880.

Butterworth, G., & Cochran, E. (1980). Towards a mechanism of joint visual attention in human infancy. *International Journal of Behavioral Development*, *3*(3), 253-272.

Chang, L. M., & Deák, G. O. (2019). Maternal discourse continuity and infants' actions organize 12‐month‐olds' language exposure during object play. *Developmental science*, *22*(3), e12770.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & language*, *25*(2), 141-168.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental psychology*, *40*(2), 271.

Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12-and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental psychology*, *36*(4), 511.

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.

Feinman, S. (1982). Social referencing in infancy. *Merrill-Palmer Quarterly (1982-)*, 445-470.

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.

Flom, R., & Pick, A. D. (2003). Verbal encouragement and joint attention in 18-month-old infants. *Infant Behavior and Development*, *26*(2), 121-134.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. Psychology Press.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological science*, *6*(5), 314-317.

Mangold, P., 2022. INTERACT User Guide. Mangold International GmbH (Ed.).

Newson, J., & Newson, E. (1975). Intersubjectivity and the transmission of culture: On the social origins of symbolic functioning. *Bulletin of the British psychological society*.

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., Triesch, J., & Deák, G. O. (2023). Variability in infant social responsiveness: Age and situational differences in attention-following. *Developmental Cognitive Neuroscience*, *63*, 101283.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244-262.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS one*, *8*(11), e79659.