Using manual actions to create visual saliency: an outside-in solution to sustained attention and joint attention

Jane Yang
jane.yang@austin.utexas.edu
University of Texas, Austin

Linda Smith
smith4@iu.edu
Indiana University

David Crandall
djcran@indiana.edu
Indiana University

Chen Yu
chen.yu@austin.utexas.edu
University of Texas, Austin

Abstract

Human cognition is shaped by our bodies and body actions. The influence of embodiment on cognition is particularly crucial during early development. Recent evidence shows that young children use bodily actions to accomplish cognitive and social tasks that may later be solved internally. In the present study, we propose that a sensorimotor mechanism to hand-eye coordination is through a full path from manual action, to visual saliency in view, and to visual attention. To provide a rigorous test of this full pathway, we analyzed multimodal behavioral data collected from parent-infant toy play. We focused on linking infants' manual actions with visual properties in the infant's view and infant attention. Further, we extended our analyses to quantify the effects of manual actions on one's own visual attention, the effects of the infant's actions on parent attention, and the effects of the parent's actions on infant attention. Our results suggest that both infants' and parents' actions in joint play create visual saliency of objects in play to support visual attention and joint attention.

Keywords: egocentric vision; sustained attention; joint attention; embodied cognition; parent-child interactions

Introduction

Infants are active learners: they explore and learn about the world by acting on it. Infants' bodily actions not only create visual data with unique properties in their first-person view but also elicit child-directed speech from responsive caregivers (Suarez-Rivera, Linn, & Tamis-LeMonda, 2022). As a result, infants learn the names of objects in their hands in both the laboratory and home environments (Suarez-Rivera et al., 2022; Yu & Smith, 2012), providing growing evidence that motor development is closely tied to early word learning (Yu & Smith, 2012; Iverson, 2010).

Among the many ways that infants' manual actions impact early learning, one critical path is hand-eye coordination – holding and looking at an object at the same time. During the second year, a developmental period with rapid cognitive and motor development, visual attention is intimately linked to developing sensorimotor abilities. One study showed that infants with high hand-eye coordination during toy play create coherent distributions of visual attention over a set of visual objects in the environment, whereas individuals with low hand-eye coordination demonstrate disrupted visual attention (Abney, Karmazyn, Smith, & Yu, 2018). By linking infants' action and gaze data during object play with the outcome of object name learning, another study found that infant visual attention alone did not predict word learning. Instead, coordinated, multimodal attention—when infants' hands and eyes were attending to the same object—predicted word learning (Schroer & Yu, 2022). A study in the home environment showed that objects of infant play elicit parent naming through which infants learned the names of objects in their hands (Suarez-Rivera et al., 2022). Hand-Eye coordination has also been studied in the context of parent-infant joint attention. Two recent studies showed that joint attention did not arise through gaze following but rather through the coordination of gaze with manual actions on objects as both infants and parents attended to their partner's object manipulations (Yu & Smith, 2013, 2017). Moreover, dyad differences in joint attention were associated with dyad differences in hand following.

Even though the importance of hand-eye coordination in infancy has been well-documented, little is known about the underlying sensorimotor processes that support hand-eye coordination. Previous research showed that infants visually attend longer to objects in their hands compared to those that are not (Ruff & Lawson, 1990; Yu & Smith, 2017; Deak, Krasno, Triesch, Lewis, & Sepeta, 2014). Further, objects of infant play are visually salient in infants' visual fields (Yu & Smith, 2012). Taken together, we propose a sensorimotor pathway to coordinating hands and eyes that begins with a manual action the object that creates a visual salience, that yields gaze directed to the object being handled. The present study aimed to provide a rigorous test of this full pathway. Towards this goal, we focused on linking infants' manual actions with visual properties in the infant view and infant attention. Further, we extended our analyses in the context of parent-child social interaction, quantifying the effects of manual actions on one's own visual attention, the effects of the infant's actions on parent attention, and the effects of the parent's actions on infant attention. Our results suggest that both infants' and parents' actions in joint play create visual saliency that supports visual attention and joint attention. Figure 1 provides an overview of multiple sensorimotor pathways examined in the present study, offering a sensorimotor pathway to hand-eye coordination between infants and parents.

Recent advances in head-camera and head-mounted eye tracking technologies have begun to uncover unique visual properties from an infant's egocentric view that are created by the body and body movements (Yu & Smith, 2012; Sullivan, Mei, Perfors, Wojcik, & Frank, 2021; Bambach, Smith,

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

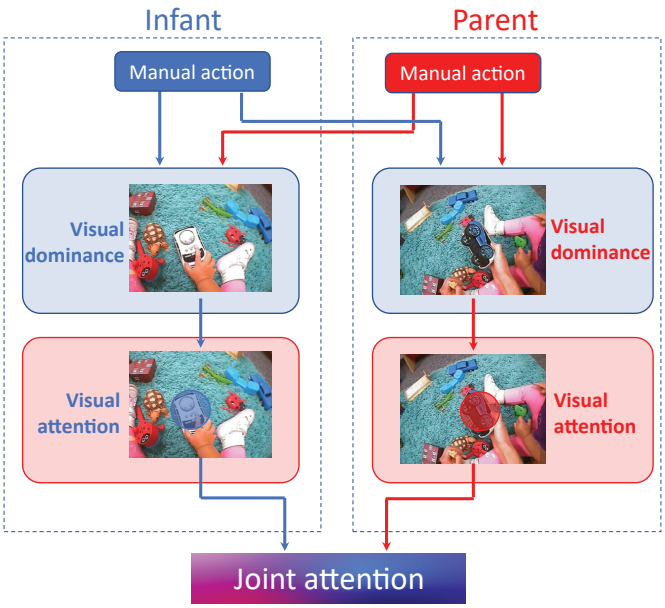

Figure 1: An overview of the sensorimotor pathways proposed and tested in the present study. Manual actions from both infants and parents create visual dominance of the held objects in both the infant's and parent's views respectively. This behaviorally-created visual saliency attracts both infant attention and parent attention to the held objects, which facilitates the establishment of joint attention between the two social partners.

Crandall, & Yu, 2016). One study shows that head stability and body posture behaviorally control the visual input in the infant's view (Méndez, Yu, & Smith, 2021). Another study characterizes the salience of objects placed by infants or parents by analyzing egocentric images right after a manual action was produced (Anderson, Seemiller, & Smith, 2022). The results from that study show that adult observers can quickly find the objects placed by either infants or parents in a visual search task. The present paper differs from the previous studies in several important ways: 1) We directly measured infants' and parent's attention in the moments of manual action through head-mounted eye tracking; 2) we measured and linked visual properties of objects in play with visual attention; 3) we measured and compared the effects of manual engagement from parents and infants in free-flowing interaction; and 4) we compared and measured the effects of manual actions on infants' attention, parents' attention, and joint attention between the two social partners.

Data

The data used in this analysis were collected from free-flowing parent-child play sessions (n = 26), each involving the same set of 24 objects. All children were between the age of 15 to 24 months (M = 19.3, SD = 2.1, *Min* = 15.2, *Max* = 24.3). Each session lasted an average of 7.01 minutes (range 3.74 - 11.69 min). Together, the data used in the present study

Figure 2: An overview of the experimental setup from the parent's egocentric (left), third-person (middle) and the infant's egocentric (right) views. The black cross-hair in the egocentric image indicates the infant's gaze point.

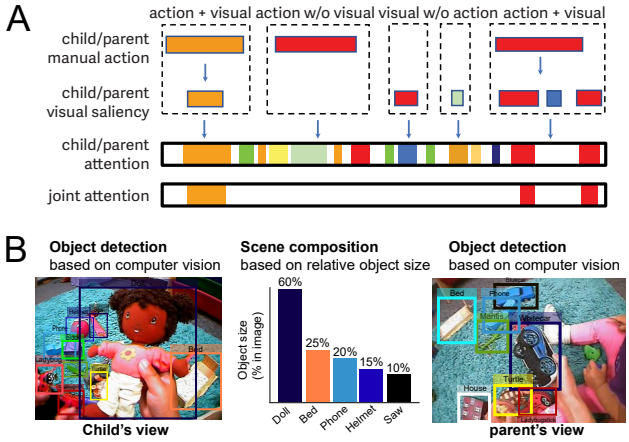

Figure 3: Panel A demonstrates three types of sensorimotor events in our analyses. Panel B provides an example of the object detection data after processing (left) and the corresponding extracted toy sizes (right).

is from 3.24 hours of interaction video in total, with 350,042 image frames each extracted from the infant's and the parent's egocentric views (30 frames per second). Figure 2-middle shows a third-person view of the experiment setup.

At the beginning of the experiment, the 24 toys were randomly spread on the floor. The parents were asked to play as they would at home and to keep their children engaged with those toys. During a play session, the parent and infant each wore a head-mounted eye-tracker with a front-facing camera capturing their egocentric view. An example of the infant's view can be seen in Figure 2-right (Franchak, Kretch, Soska, Babcock, & Adolph, 2010; Yu & Smith, 2013). The eye camera was mounted on the head and pointed to the right eye of the participant. The scene camera captured the first-person view from the participant's perspective, with a 90° horizontal field. Each eye tracking system recorded both the egocentric view video and gaze direction in that view, with a sampling rate of 30 Hz. Three third-person view cameras were also used to record the play session from a distance. A detailed description of this study can be found in (Yu, Zhang, Slone, & Smith, 2021).

Data processing

Egocentric videos, eye videos, and third-person view videos were first synchronized in time and decomposed into image frames. We then followed a calibration procedure commonly

used in head-mounted eye tracking (the details provided in (Yu et al., 2021)). After calibration, a cross-hair was superimposed in each of the egocentric images to indicate the wearer's visual attention in view. From calibrated videos, we annotated four behaviors: infant gaze, parent gaze, infant manual action, and parent manual action. We also used computer vision algorithms to automatically detect the 24 objects in each of the infant's and parent's image frames. Based on the relative sizes of objects in view, we operationally defined visual dominance (described in a later section).

Gaze data Each of the 24 toys in toy play sessions was identified as a region-of-interest (ROI). Coders watched the calibrated egocentric videos frame-by-frame and coded an ROI for each of the frames using an in-house program. An example of calibrated egocentric view with crosshair is shown in Figure 2-left. In total, 78.4% of frames from infant's view and 80.8% of frames from parent's view contain ROIs to toy objects (274,539 frames of infant's gaze data and 282,857 of parent's gaze data in total).

Joint attention Infant and parent gaze data were aligned in time and compared at the frame level to find moments of joint attention (JA). For every frame, JA was objectively defined as when parent and infant were gazing at the same object – no other behaviors were needed to count as JA. For the analyses, a bout of JA had to last at least 500ms but could include short looks away from the attended object (Yu & Smith, 2017) if gaze switched back to the jointly attended object.

Manual actions on objects Coders watched a play session from the views of multiple cameras and annotated, frame-by-frame, the object with which the infant's or parent's hands made contact. For each of the two social partners, coders went through the session twice, once to annotate manual action from the left hand, and then to annotate the right hand. In total, there were 1,234 instances of infants' manual action events (M = 47.46, SD = 30.15) and 1,446 instances of parents' manual action events (M = 55.62, SD = 25.76).

Object size as visual saliency Visual size is a well-documented property of visual salience that attracts people's attention (Proulx & Green, 2011). A growing literature using infant head-mounted cameras shows that infants move their bodies and thus change the visual size of objects in their first-person view. Accordingly, we used the visual size of objects in an egocentric view as a measure of visual salience of those objects created by manual actions. To do this, as shown in Figure 3b, we used a pre-trained deep learning model (YOLOv3) to automatically detect individual objects in view (Redmon & Farhadi, 2018). The trained model provided up to 24 bounding boxes per frame, indicating the location of each of the 24 toys in view. We then computed the visual size of an object by calculating the fraction of the area of the frame covered by the area of the bounding box, shown in Figure 3b-middle.

Visual dominance Using object size, we defined an object in view as visually dominant if its bounding box occupied at least 5% of the area of the frame and it was at least twice as big as the second-largest object within the same frame. Based on the two criteria, we identified a dominant object in each frame when applicable. In Figure 3B-left, the toy doll is greater than 5% of the view and is also at least twice as big as the second largest object (toy bed in the back), so the toy doll is defined as the visual dominant object of this frame. In Figure 3B-right, the toy car is not a visual dominant object of the frame because its size is comparable to the toy bed in the back, so there is no visual dominant object in that frame.

Statistical analyses. Analyses were conducted at the corpus level. The independent measures were the visual size of objects manipulated either by infants, by parents, or jointly by both. The dependent measures were frame-by-frame measures of visual attention and joint attention. Mixed-effect linear regression models were conducted using the lme4 package in R (Version 3.6.1; (Bates, 2010)). Individual infants and the specific toy objects were random variables in all analyses.

Results

We first report the results on the sensorimotor pathways that lead to infant attention (the left side in Figure 1), followed by the sensorimotor pathways to parent attention (the right side in Figure 1). Then we focus on the joint pathways to joint attention.

Pathways from manual action, to visual saliency, to infant attention

We first examined a full pathway from infant manual action, to visual saliency of the held object, and to infant attention. To do this, we identified manual action events and calculated the mean size of the held object in each event. Figure 4A shows a histogram of object size across individual action events. Next, we used a median split on object size to divide action events into two groups: 1) with a larger size and 2) with a smaller size. The held object was much bigger and therefore more visually salient in the larger group (M = 19.1%) than the held objects in the smaller group (M = 4.4%). We next compared infant visual attention in the two event groups and found that when a manual action on an object created visual salience of that object, the infant attended to the salient object much more often than the held object that was not salient ($M_{salient}$ = 0.641; $M_{not\_salient}$ = 0.281; β = 2.086, SE = 0.115, $p < 0.0001$).

Next, we examined a parallel pathway from parent manual action, to visual saliency of the held object, and to infant attention. Similar to the data analytics approach above, we categorized the events of parent manual action into two groups based on a median split on the held object size: action events that created a larger size of the held object (M = 13.8%) and action events that created a smaller size of the held object (M = 2.5%). A comparison of infant attention within the two action types revealed that when a manual action from parents

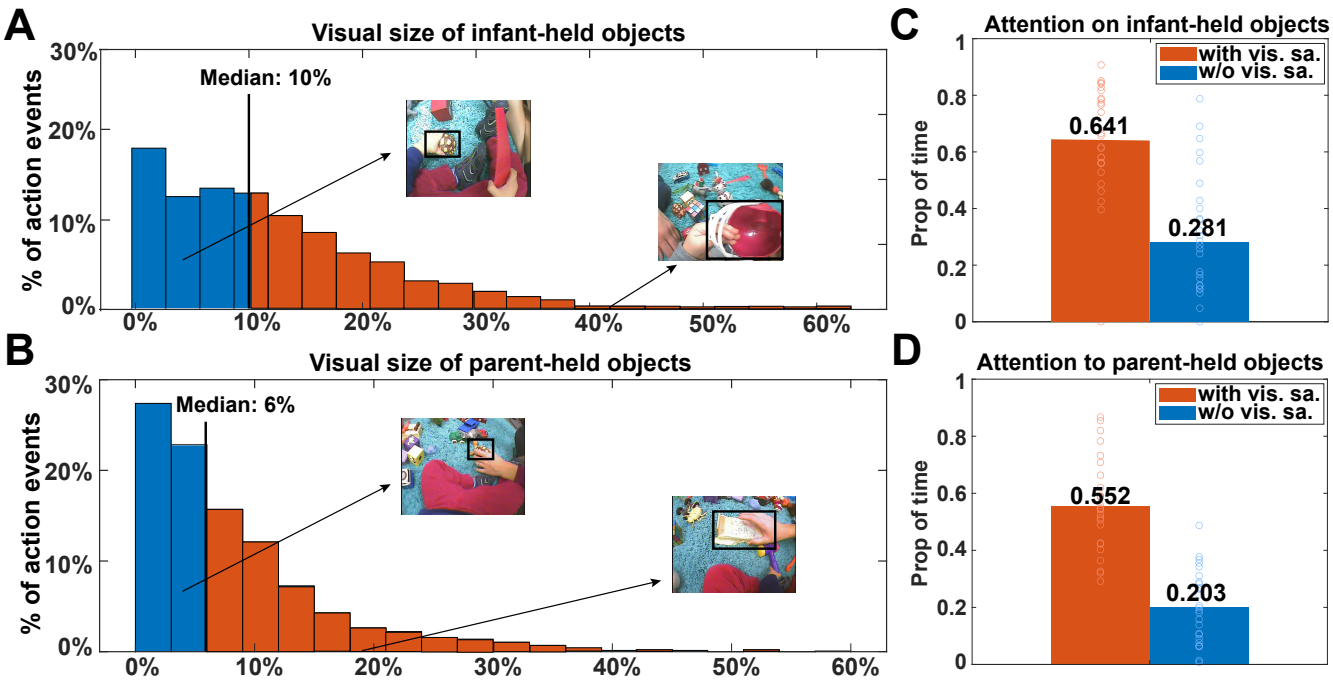

Figure 4: Plots A and B display the distribution of the average size of infant/parent-held objects from the infant's egocentric view. The plots each contain smaller pictures representing an instance of a salient object (has size above median) and a not salient object (has size below median). Plots C and D show the proportion of time the infant attended to salient and not salient infant/parent-held objects.

created more visual salience of the held object, the infant attended to the target object more than the held object which was smaller in size and therefore less visually salient. Thus, parents can use their manual actions to attract the infant's attention by creating visual salience of the held object in the infant's view.

We next conducted a head-to-head comparison of the effects of infant and parent manual actions. In terms of the size of a held object, the objects held by infant itself were larger in view and therefore more visually salient (see the distribution in Figure 4A), compared with the objects held by parents (see the distribution in Figure 4B). However, in terms of the effects on the infant's attention, parent manual actions are as effective as infant manual actions ($M_{salient\_parent\_action}$ = 0.463; $M_{salient\_infant\_action}$ = 0.538; β = 0.026, SE = 0.030, p = 0.263107). When a held object was relatively salient in view, either through an infant's or parent's manual action, the infant attended longer to the held object relative to the held object without visual saliency.

One plausible interpretation of the results reported above is that visual saliency attracts infant attention independent of manual action. To test this possibility, we zoomed into visual dominance events, which are operationally defined as an object visually dominant in view, and categorized those events into four types: 1) with infant manual action, when the infant held the visual object over 50% of time within a dominance event; 2) with parent manual action, when the parent held the visual object over 50% of time; 3) with joint action

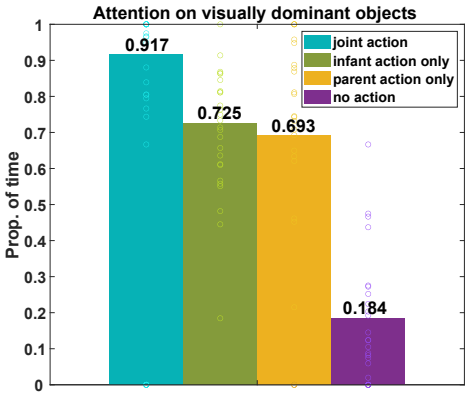

Figure 5: The proportion of time the infant attended to the visually dominant object when it is held by infant itself, parent, both, or neither.

from both infant and parent; and 4) no manual action. As shown in Figure 5, when the visual dominance of an object was created by either infant action, parent action, or joint action, the infant attended to the object more during those moments than the moments when visual dominance of an object was not created by any manual action (M_{joint_action} = 0.917, $M_{infant\_action\_only}$ = 0.725, $M_{parent\_action\_only}$ = 0.693, M_{no_action} = 0.184). To determine whether the proportion of time of manual action can be considered as a predictor for infant attention, we selected three subsets of data and constructed a simple linear model to fit each subset. We grouped instances where dominance was created by joint action and those that

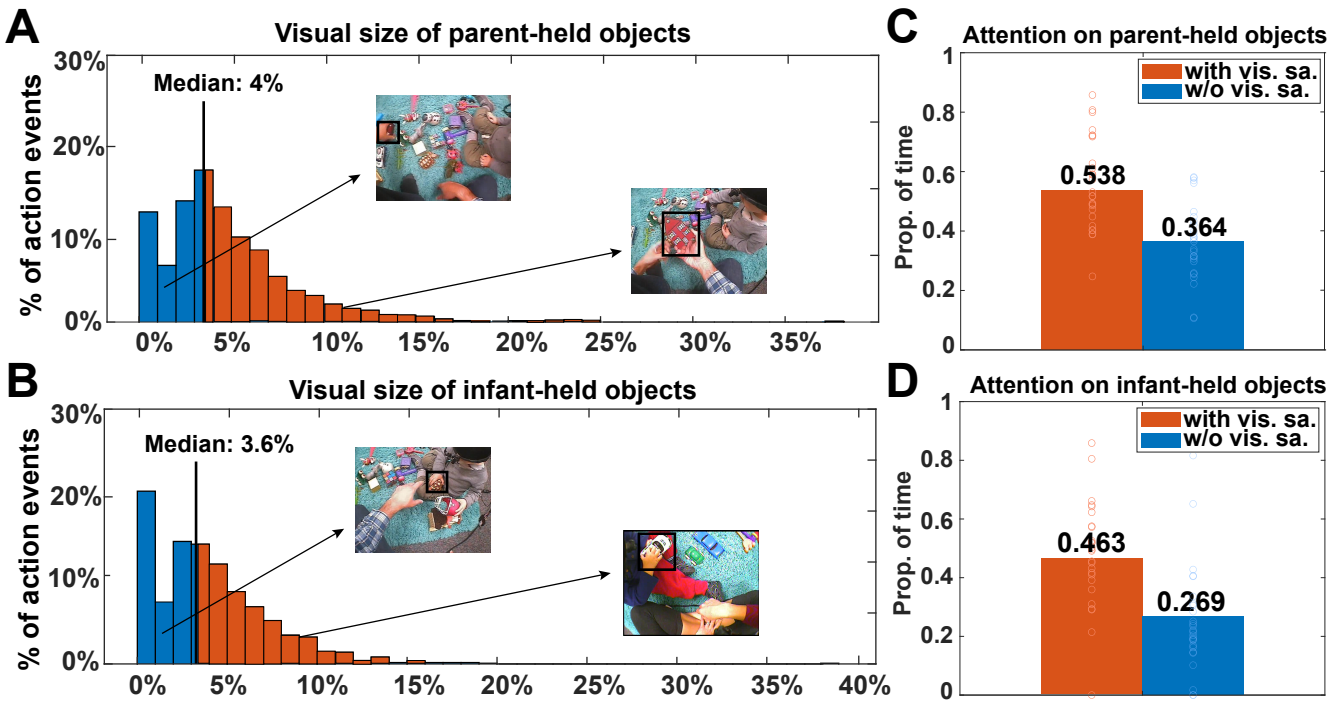

Figure 6: Plots A and B display the distribution of the average size of parent/infant-held objects from parent's egocentric view. A and B each contain smaller pictures representing an instance of salient object (has size above median) and a not salient object (has size below median). Plots C and D show the proportion of time parent attended to salient and not salient parent/infant-held objects.

were created without action, and we found that the effect of proportion of time of holding is significant ($\beta = 0.608$, $SE = 0.055$, $p < 0.0001$). The result was comparable when we fit the model on instances where dominance was created by infant's manual action and without action ($\beta = 0.341$, $SE = 0.026$, $p < 0.0001$). Lastly, the result was also significant when we compared parent action instances to no action instances ($\beta = 0.342$, $SE = 0.032$, $p < 0.0001$).

Taken together, the results suggest that the visual dominance of an object on its own did not attract the infant's attention. Only when the object's visual saliency was created by manual actions, infants attended to those held objects more than the held objects without visual saliency.

Pathways from manual action, to visual saliency, to parent attention

Using the same approach described in the previous subsection, we examined the full pathways from manual action, to visual saliency in the parent's view, and to parent attention. Figure 6-left shows the histograms of the size of objects, either held by infants or by parents. The shape of the histograms in the parent's view is similar to the overall shape of the histograms in the infant's view. However, consistent with previous findings (Yu & Smith, 2012), objects appear significantly smaller in the parent's view compared with object appearance in the infant's view ($M_{infant\_held\_obj\_parent\_view} = 4.0\%$, $M_{parent\_held\_obj\_parent\_view} = 4.8\%$, $M_{infant\_held\_obj\_infant\_view} = 11.8\%$, $M_{parent\_held\_obj\_infant\_view} = 8.1\%$; $t = 28.3$, $df = 3553.2$, $p < 0.0001$). The recent literature on infant egocentric vision suggests that the difference is primarily due to different body sizes and arm lengths between adults and infants, which created different distances between objects in play and the head-mounted camera. Manual action events from infants and parents are divided respectively into two groups using a median split from the corresponding distribution. For parent actions shown in Figure 6C, the parent attended more to a held object if that object was larger in view ($M_{salient} = 0.552$, $M_{not\_salient} = 0.203$; $\beta = 2.500$, $SE = 0.315$, $p < 0.0001$). The pattern is consistently observed in the infant's manual actions ($\beta = 2.086$, $SE = 0.115$, $p < 0.0001$). Thus, just like infant attention, parent attention is equally likely to be influenced by the visual saliency of objects created by either infants' or parents' manual actions ($M_{salient} = 0.538$, $M_{not\_salient} = 0.364$; $\beta = 2.292$, $SE = 0.316$, $p < 0.0001$). We also compared visual dominance events with or without manual actions. Figure 7 shows that parent attention was not attracted by the visual saliency alone when that saliency was not created by either infants' or parents' manual actions ($M_{joint\_action} = 0.791$, $M_{infant\_action\_only} = 0.666$, $M_{parent\_action\_only} = 0.486$, $M_{no\_action} = 0.026$). When we compared instances where dominance was created by joint action to instances created without action, the result is significant ($\beta = 0.558$, $SE = 0.028$, $p < 0.0001$). The result is likewise significant when we com-

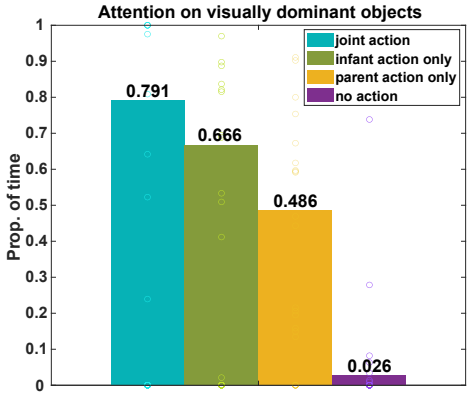

Figure 7: The proportion of time parent attended to visually dominant object when it's held by either infant or parent, both, or neither.

pared infant action instances to no action instances ($\beta = 0.364$, $SE = 0.029$, $p < 0.0001$). Lastly, the result is also significant when we compared parent action instances to no action instances ($\beta = 0.368$, $SE_{parent\_action} = 0.036$, $p < 0.0001$).

Pathways from manual action, to visual saliency, and to joint attention

In the last set of analyses, we examined the effects of infants' and parents' manual actions on joint attention. For infant manual actions, we identified the moments that an action event created visual dominance in the infant's view and in the parent's view respectively. Combining the data on visual dominance from the two views, we categorized infant action events into four categories: 1) with visual dominance in the infant's view; 2) with visual dominance in the parent's view; 3) visual dominance in both views; and 4) no visual dominance in either view. We next measured joint attention in those types of infant action events, shown in Figure 8A, and found that only when an infant's manual action simultaneously created visual dominance in both the infant's and the parent's views, the dyad was likely to jointly attend to that object ($M_{joint\_vis\_dom} = 0.451$, $M_{infant\_vis\_dom} = 0.287$, $M_{parent\_vis\_dom} = 0.205$, $M_{no\_vis\_dom} = 0.131$). We constructed three linear mixed-effect models and performed a likelihood ratio test to quantify the effect of infant and parent visual dominance in predicting joint attention. We found that considering infant and parent's visual dominance jointly as fixed effects significantly better predicted the proportion of joint attention (joint dominance compares to infant dominance only: $\chi^2(1) = 20.229$, $p < 0.0001$, joint dominance compares to parent dominance only: $\chi^2(1) = 109.37$, $p < 0.0001$). The same analysis was conducted on parent manual actions. Figure 8B shows the effects on joint attention from parent manual actions are comparable with those from infant manual actions ($M_{joint\_vis\_dom} = 0.424$; $M_{infant\_vis\_dom} = 0.411$; $M_{parent\_vis\_dom} = 0.160$, $M_{no\_vis\_dom} = 0.087$; joint dominance compares to infant dominance only: $\chi^2(1) = 3.857$, $p = 0.04954$, joint dominance compares to parent dominance only: $\chi^2(1) = 158.44$, $p < 0.0001$).

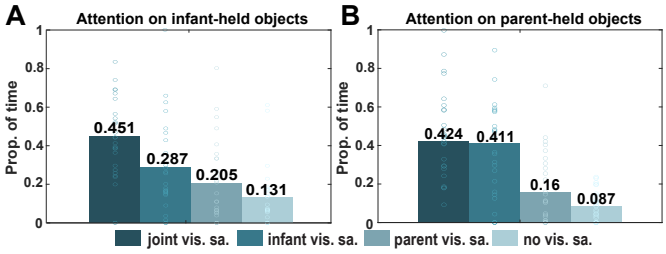

Figure 8: Plot A shows the proportion of time of joint attention on infant-held objects when holding created saliency in infant's, parent's, both agents' views, or neither. Plot B shows comparable results for parent-held objects.

General Discussion

As shown in Figure 1, the present study reveals multiple sensorimotor pathways that link manual action, visual saliency, and visual attention. In the context of parent-infant free-flowing social interaction, the effects of parents' actions on infant attention are as potent as the effects of infants' attention on their own attention, and likewise for infants' and parents' actions on parents' attention. In all of the pathways, manual action affects visual attention by creating visual saliency in view. Both infants' and parents' actions in joint play create visual saliency of objects in play that supports visual attention and joint attention.

These results provide new empirical evidence on dependencies among manual action, visual saliency, and visual attention in everyday parent-child social interactions. The results also highlight bidirectional influences between manual action and visual attention, both within the infant's own cognitive system and between the infant and the social partner. Vision and visual attention provide sensory information that is needed to guide manual action. Meanwhile, object handling has direct effects on the visual input that directly influences gaze direction.

Those results support the theoretical view of human development as a multicausal system with many pathways (Kelso, 1995; Thelen, Kelso, & Fogel, 1987). Even though a multi-pathway system is complex, the inherent redundancy in such a system offers robustness in developmental outcomes. Given multiple pathways to influence infants' attention and parents' attention to achieve joint attention, those functionally redundant pathways can compensate for one another, providing alternative routes at the system level to recover from single pathway failure. This unique property of a multi-pathway system also has practical relevance as exploiting these pathways may offer opportunities for effective interventions for atypical developing populations. They may not have to be trained to use the same pathways as typically developing children, given some pathways may not be feasible. Instead, they can just exploit alternative pathways that are feasible for them to reach the same function end.

References

Abney, D. H., Karmazyn, H., Smith, L. B., & Yu, C. (2018).

Hand-eye coordination and visual attention in infancy. In *Cogsci.*

Anderson, E. M., Seemiller, E. S., & Smith, L. B. (2022). Scene saliencies in egocentric vision and their creation by parents and infants. *Cognition*, *229*, 105256.

Bambach, S., Smith, L. B., Crandall, D. J., & Yu, C. (2016). Objects in the center: How the infant's body constrains infant scenes. In *2016 joint ieee international conference on development and learning and epigenetic robotics (icdl-epirob)* (pp. 132–137).

Bates, D. M. (2010). *lme4: Mixed-effects modeling with r*. Springer New York.

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.

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., & Adolph, K. E. (2010). Head-mounted eye-tracking of infants' natural interactions: a new method. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 21–27).

Iverson, J. M. (2010). Developing language in a developing body: the relationship between motor development and language development. *Journal of Child Language*, *37*(2), 229–261. doi: 10.1017/S0305000909990432

Kelso, J. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT press.

Méndez, A. H., Yu, C., & Smith, L. B. (2021). One-year old infants control bottom-up saliencies to purposely sustain attention.

Proulx, M. J., & Green, M. (2011, 11). Does apparent size capture attention in visual search? Evidence from the Müller–Lyer illusion. *Journal of Vision*, *11*(13), 21-21. Retrieved from https://doi.org/10.1167/11.13.21 doi: 10.1167/11.13.21

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental psychology*, *26*(1), 85.

Schroer, S. E., & Yu, C. (2022). Looking is not enough: Multimodal attention supports the real-time learning of new words. *Developmental Science*, e13290.

Suarez-Rivera, C., Linn, E., & Tamis-LeMonda, C. S. (2022). From play to language: Infants' actions on objects cascade to word learning. *Language Learning*, *72*(4), 1092–1127.

Sullivan, J., Mei, M., Perfors, A., Wojcik, E., & Frank, M. C. (2021, 05). SAYCam: A Large, Longitudinal Audiovisual Dataset Recorded From the Infant's Perspective. *Open Mind*, *5*, 20-29. Retrieved from https://doi.org/10.1162/opmi_a_00039 doi: 10.1162/opmi$_{a0}$0039

Thelen, E., Kelso, J. S., & Fogel, A. (1987). Self-organizing systems and infant motor development. *Developmental review*, *7*(1), 39–65.

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.

Yu, C., & Smith, L. B. (2017). Hand–eye coordination predicts joint attention. *Child development*, *88*(6), 2060–2078.

Yu, C., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. *Proceedings of the National Academy of Sciences*, *118*(52), e2107019118.