Published in final edited form as:

Vis cogn. 2008 January 1; 16(4): 430–438.

Preschoolers' oculomotor behaviour during their observation of an action task

James D. Morgante, Jeffrey M. Haddad, and Rachel Keen

University of Massachusetts, Amherst, MA, USA

Abstract

Adult and child oculomotor behaviour was measured during the observation of a naturalistic action task. Adult and 4-year-old participants viewed a video presentation of an actor making a root-beer float. Eye movements were monitored to examine fixation patterns during a series of means–end action sequences. Look-ahead fixations (i.e., gaze anticipation) were measured to determine if children would saccade to the goal site prior to the completion of each action sequence as adults do. Look-ahead fixations were observed in both the adult and child groups, with no significant age differences found in levels of gaze anticipation.

The ability to interpret goal-directed sequential actions, as well as perform multiple steps in problem-solving tasks in a planful manner, has obvious adaptive significance. As the building blocks of social cognition, human beings need to comprehend these types of behaviours in order to understand desires, intentions, and agency of others. The capacity to link actions that occur together enables one to form coherent representations of action sequences at the level of intention as opposed to the level of the action unit. Adults and children show the propensity to interpret others' actions as well as perform these actions (Lillard & Flavell, 1990; Sommerville & Woodward, 2005; Trabasso, Stein, Rodkin, Munger, & Baugh, 1992; Willatts, 1999). In this study, we examined eye fixations of adults and 4-year-olds during the observation of a goal-directed daily living task in order to gain further insight into movement planning.

Explorations of oculomotor behaviour during object manipulation and naturalistic action tasks have revealed “look-ahead” fixations in which adults make a saccade to an object's spatial location prior to the execution of an action sequence on that object. Look-ahead fixations have been observed in tea making (Land, Mennie, & Rusted, 1999), hand-washing (Pelz & Canosa, 2001), and sandwich-making tasks (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003). With each of these tasks, eye and hand movements were measured as participants completed an activity of daily living. As a component of natural behaviours, fixation ahead of the action on an object rather than on the action itself shows awareness of movement planning. Hayhoe et al. (2003) interpreted these fixations as indicative of adults acquiring the spatial location of an object to guide the next movement or to plan a reach. When localizing an object in the tea-making task, Land et al. (1999) reported several instances where prior fixation on an object to be picked up shortly after rendered the object more readily found.

Beyond these naturalistic tasks, similar fixations toward the location of a future action were observed by Johansson, Westling, Bäckström, and Flanagan (2001) in a controlled experimental situation where adults reached for and grasped a bar, moved it past an obstacle, and used it to press a target switch. In a follow-up investigation that compared observation of another's action with an eye–hand coordination task performed by the participant, Flanagan

and Johansson (2003) observed that during a block-stacking task, gaze was predictive rather than reactive during both self-produced and observed action. Eye arrival at a contact zone prior to the hand was observed under both conditions. Together, these findings indicate that look-ahead fixations are an integral component for motor planning and the interpretation of action.

Very young children appear to be capable of understanding actions performed by others. Using the visual habituation paradigm, researchers have sought to determine when infants are able to interpret simple actions of others (i.e., reaching and grasping) as well as relate sequential actions to overarching goals. As early as 6 months, infants show the propensity to interpret the simple action of grasping as being goal directed (Woodward, 1998). The ability to comprehend causal relations, such as those involved in executing means–ends actions, has been observed in infants from 5½ (Baillargeon, Graber, DeVos, & Black, 1990) to 12 months of age (Woodward & Sommerville, 2000), when infants view others' actions rather than acting themselves.

Extending the findings from visual habituation studies that measure preferential looking, Falck-Ytter, Gredeback, and von Hofsten (2006) utilized a gaze-recording technique to examine oculomotor behaviours during a simple goal-directed action. The eye movements of adults, 6-, and 12-month-olds were compared in the context of a video presentation in which a toy was moved into a bucket. Consistent with findings from the aforementioned eye–hand coordination tasks, adults and 12-month-olds predicted the actor's intention by shifting their fixation to the goal of the action (i.e., the bucket) prior to the arrival of the hand. In comparison, 6-month-old infants did not show look-ahead fixations but instead shifted fixation to the goal state after the hand had arrived. The authors concluded that infants as young as 12 months show patterns of fixation that are comparable to adults when observing another person's goal-directed action.

The goal of the present investigation was to examine the look-ahead fixation phenomena in an observation task with an action sequence comparable to naturalistic tasks used in the visual memory and motor planning literature. In contrast, the action event observed by participants in the study by Falck-Ytter et al. (2006) was presented in a visual habituation format that was both simple and repetitive. Across trials, children and adults viewed the same action movement enacted on comparable objects. After viewing the first trial, look-ahead fixations may have been reactive as opposed to predictive as a result of the repetition in the event sequence.

To broaden the scope of current findings reported for a simple action (i.e., reaching and grasping), we modelled our task after a naturalistic food preparation task that children might observe every day. Adults and 4-year-old participants were asked to watch a video presentation of an actor making a composite confection (i.e., root-beer float). Preschool-aged children were chosen for this task, as it is the age of acquisition for adaptive skills such as preparing and serving food (Bricker & Pretti-Frontczak, 2000). A naturalistic task involving food preparation is appropriate for 4-year-olds, since children of this age begin to pour liquids and prepare food for eating by removing the inedible parts (Bricker & Pretti-Frontczak, 2000). The focus of our investigation was to compare child and adult fixations during several sequential events to determine if (a) the actors' actions were anticipated rather than followed by the viewer when the deployment of a series of actions is necessary to achieve an overarching goal, and (b) young children show similar look-ahead fixation patterns compared to adults when there is greater complexity in the causal relations of the intentional action sequence.

METHOD

Participants

There were two groups of participants, with a total of 22 participants in each group. The adult group consisted of 9 male and 13 female adults (range = 18–27 years). The 4-year-old group

consisted of 13 male and 9 female children (range = 3 years; 10 months to 4 years; 2 months). Thirteen additional participants, 4 adults and 9 children, were tested and removed from the final sample because of unsuccessful calibration ($n = 7$) or failure to obtain reliable eye data ($n = 6$). Participants had normal or corrected-to-normal vision (with soft contact lenses).

Several sources, including birth records from local town halls, birth listings in the newspaper, and a commercial source were used to obtain the names of children for this study. Guardians were initially contacted via letter followed up by a phone call. Participants from the adult group were recruited from the student body of the University of Massachusetts at Amherst.

Apparatus

The video was presented to participants on a 69 cm television screen. Participants were asked to sit in a chair approximately 1 m away from the screen. While adults sat independently, child participants sat in the lap of their guardian. The television was mounted on a platform 84 cm off the ground so that the eye-tracking system could be positioned in the space below. A corneal reflection eye-tracking system in association with a head-tracking system (Applied Science Laboratories Model 504) was used to track the participant's eye movements at 60 Hz. The eye tracker monitored the right eye, and viewing was binocular. The spatial resolution of the eye tracker was approximately 0.5 degrees of visual angle.

Procedure

Experimental sessions lasted between 15 and 30 min. Each session consisted of the following events. The participant sat in front of the television screen and eye-tracking camera. In order to calibrate the point of gaze, static objects were presented throughout the visual scene on the television screen. Various animated objects (e.g., aeroplane, butterfly, car) appeared sequentially at nine points in a 3×3 grid that covered the visual field. If the point of gaze was not exactly on the object, a linear offset was performed to correct it. After each calibration, point of gaze accuracy was tested by having each participant watch a display of static objects and dynamic objects that travelled across the visual field at known points in space. If the coordinates of the point of gaze did not correspond to the objects in the known positions during the calibration test, the data from that participant was not included in the final sample. The same calibration was used for both child and adult participants. After the calibration, a short video of an actor producing an action sequence was played.

Participants were instructed that the video they would watch was of a person making a snack. In the video, the actor prepared a root-beer float using soda, ice cream, sprinkles, and a cherry. The video was created in a style similar to one of a food preparation television show. At the beginning of the video, the actor introduced herself, described what she would be making, and showed the viewer a finished product of her snack. After showing the finished product, the actor showed the viewer each of the ingredients she would be using to make her snack in the order the item was to be used. She did this by picking the item up off of the table in front of her and labelling it for the viewer.

Once all of the ingredients had been identified, the camera cut to a view of the table and the actor's midsection, shoulders to waistline, and the root-beer float was made. To create the root-beer float, the actor performed five intentional actions. These actions included pouring soda, adding two scoops of ice cream, shaking sprinkles, and topping the float off with a cherry. Ingredients were introduced one at a time on the right side of the screen while the glass remained stationary on the left. The actor utilized a placement mark, a small "X" on the tablecloth visible only to the actor, to ensure equidistant placement for each of the items. The starting position of each ingredient was no greater in proximity to the glass than the other items. This camera shot was used for two reasons: (1) to eliminate referential communication cues (i.e., eye gaze)

on the part of the actor that could potentially contribute to particular observer responses, and (2) to create a viewing condition in which eye movements could be measured with greater confidence.

The portion of the video with the five intentional actions was shown twice, for a total of ten action events to be scored. For each action, onset and completion times were established. The action onset was used to refer to the moment in which one of the intentional action sequences began. Action onset was defined as the moment in the event when the actor's goal was articulated through body movement. This movement was determined for each event and labelled as a specific frame in the sequence. Action completion time for each event was defined as the moment at which the necessary steps to complete the task had been completed and the event was about to end (i.e., the item was about to be added to the glass). The video frame for the onset and completion time for each event was determined using the behavioural parameters described in Table 1.

Scoring

Participant's anticipatory behaviours were assessed by coding point-of-gaze data. Specifically, for each action, the time of eye arrival at the goal site (i.e., the glass) and time of action completion were determined (see Table 1). Anticipation was then scored using both dichotomous and continuous variables. For the dichotomous variables, participants' behaviours were determined to be anticipatory if the eye arrival time preceded the action completion time. Eye arrival at any time during this event window was scored as 1. Fixations that occurred upon action completion or afterwards were given a score of zero. In the continuous variables, the exact magnitude of the anticipation time was determined by subtracting eye arrival time from action completion time.

An overall anticipation score was calculated by totalling the number of anticipations and dividing by the number of observed action events. Events where the eye image was lost or had dropped out (e.g., excessive head turn, blink, or system error) were not included in the calculation. A minimum of three action events for each viewing was required for a participant to be included in the final analysis. Three proportion scores were calculated for each participant, one for the first viewing (five events), one for the second viewing (five events), and one for overall viewing (ten events). Independent observers were trained to score action onset, action completion, and eye arrival times using the established behavioural parameters for each event. The observers recorded the corresponding time stamp for each of these three measures and interobserver agreement was reached if the determined time was ± 1 frame (i.e., 33 ms).

The primary and secondary scorer scored all the data with interobserver reliability of 96%, 98%, and 95% agreement for action onset, action completion, and eye arrival time, respectively. One trial (out of 220) for the adult group could not be coded due to a system error ($n=1$). Eighteen trials (out of 220) for the child group could not be coded due to loss of the eye image (e.g., excessive head turn; $n=18$). The average number of scored trials for each adult and child was 9.9 and 9.2, respectively. All participants showed some anticipation.

RESULTS

The mean proportion of total trials in which adults anticipated the actor's goal prior to the completion of the action was $M=0.68$ ($SD=0.26$) for the first viewing, $M=0.74$ ($SD=0.25$) for the second viewing, and $M=0.71$ ($SD=0.22$) for overall viewing. The mean proportion of trials for children's anticipation was $M=0.58$ ($SD=0.24$) for the first viewing, $M=0.64$ ($SD=0.29$) for the second viewing, and $M=0.62$ ($SD=0.20$) for overall viewing. An analysis of variance (ANOVA) did not reveal any main effects or a viewing by group interaction. Mean proportion scores show greater anticipation for both groups during the second viewing, but this was not

significant. To test whether anticipation scores differed from what might be expected due to chance, participants' overall viewing anticipation scores were compared to 0.50, which represents anticipation and no anticipation equally often over trials. A paired-sample *t*-test for adults and children showed significant anticipation, $t(21)=4.51$, $p<.001$, and $t(21)=2.67$, $p=.014$, respectively.

An analysis of anticipation latencies was conducted to determine if there was an overall difference in degree of anticipation between the two groups. Using all trials, average anticipation latency was calculated for each participant by adding the number of milliseconds of anticipation for each event across the total number of viewed actions and dividing by the number of actions. Scores were compounded over viewings because initial analyses did not reveal differences over the two viewings. The mean duration of all events was approximately 1 s. The mean latency in which adults anticipated the actor's goal prior to the completion of the action across all trials was $M=344$ ms ($SD=160$) and $M=294$ ms ($SD=114$) for the children. A *t*-test did not reveal any group difference, $t(42)=1.2$, $p=.236$.

DISCUSSION

This study was designed to compare child and adult eye-gaze anticipation of means–end sequential events to determine if observing another's action was comparable to that of self-engagement in a series of actions. Findings from both groups indicate that, just as in self-produced actions (Hayhoe et al., 2003; Johansson et al., 2001), fixation is ahead of the action rather than on the action movement itself. In other words, both groups showed anticipatory eye movements during the observation of action sequences produced by an intentional agent. The data from our adults agrees with that of Flanagan and Johansson (2003) and extends the developmental findings of Falck-Ytter et al. (2006) to more complex sequential actions. Our preschool-aged participants did not differ from adults, but further research with younger age groups is necessary to determine when visual anticipation of a series of actions begins. The observations of Falck-Ytter et al. (2006) suggest that anticipatory oculomotor behaviours for simple actions develop between the ages of 6 and 12 months. By choosing age appropriate action sequences for very young children (i.e., under the age of 2), one could explore the nature of both simple and complex action sequences to determine the ages at which such actions are understood on the level of overarching goals (i.e., the end) as opposed to individual action units (i.e., the means).

When considering the research on media literacy and sustained attention, studies have demonstrated that young children show patterns of attention that are remarkably similar to those of older children and adults (Richards & Anderson, 2004). Because many television programmes and videos made for children are now promoted as being educational, we need to define and substantiate the nature of programmes labelled as educational for children. Measurement of eye movements could be used to evaluate learning in these videos. Visual anticipation of complex sequential actions can indicate whether a child comprehends the cognitive content of a programme. Using predictive looking as a measure of cognitive understanding, one could determine whether or not learning was occurring across viewings using gaze anticipation of upcoming events embedded in various action sequences. The ability to interpret the intention behind an action allows for enhanced cognitive organization. Our study forms the basis for future exploration of how visual-motor behaviours are tied to cognitive development through the study of additional age groups and across more contexts.

Acknowledgements

Support for this project was provided in part by National Institute of Health Grant No. R37 HD27714 to RK (formerly Rachel K. Clifton), a Public Health Service Training Grant No. 5-T32-MH16745 traineeship to JM, and a National

Institute of Health Predoctoral Fellowship No. 5F31NS050930 to JH. JH is now in the Department of Health & Kinesiology at Purdue University. RK is now in the Department of Psychology at the University of Virginia.

REFERENCES

- Baillargeon R, Graber M, DeVos J, Black JC. Why do young infants fail to search for hidden objects? *Cognition* 1990;36:255–284. [PubMed: 2265529]
- Bricker, D.; Pretti-Fontczak, K. Adaptive domain.. In: Bricker, D.; Pretti-Fontczak, K., editors. *AEPS Measurement for three to six years*. 3. Paul H. Brookes Publishing Co.; Baltimore: 2000. p. 95-114.
- Falck-Ytter T, Gredeback G, von Hofsten C. Infants predict other people's action goals. *Nature Neuroscience* 2006;9:878–879.
- Flanagan JR, Johansson RS. Action plans used in action observation. *Nature* 2003;424:769–771. [PubMed: 12917683]
- Hayhoe MM, Shrivastava A, Mruczek R, Pelz JB. Visual memory and motor planning in a natural task. *Journal of Vision* 2003;3:49–63. [PubMed: 12678625]
- Johansson R, Westling G, Bäckström A, Flanagan J. Eye-hand coordination in object manipulation. *Journal of Neuroscience* 2001;21:6917–6932. [PubMed: 11517279]
- Land M, Mennie N, Rusted J. Eye movements and the roles of vision in activities of daily living: Making a cup of tea. *Perception* 1999;28:1311–1328. [PubMed: 10755142]
- Lillard AS, Flavell JH. Young children's preference for mental state versus behavioral descriptions of human action. *Child Development* 1990;61:731–741. [PubMed: 2364748]
- Pelz JB, Canosa R. Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research* 2001;41:3587–3596. [PubMed: 11718797]
- Richards, JE.; Anderson, DR. Attentional inertia in children's extended looking at television.. In: Kail, RV., editor. *Advances in child development and behavior*. 32. Elsevier Academic Press; San Diego, CA: 2004. p. 163-212.
- Sommerville JA, Woodward AL. Pulling out the intentional structure of action: The relation between action processing and action production in infancy. *Cognition* 2005;95:1–30. [PubMed: 15629472]
- Trabasso T, Stein NL, Rodkin PC, Munger MP, Baugh CR. Knowledge of goals and plans in the on-line narration of events. *Cognitive Development* 1992;7:133–170.
- Willatts P. Development of means-end behavior in young infants: Pulling a support to retrieve a distant object. *Developmental Psychology* 1999;35:651–667. [PubMed: 10380857]
- Woodward AL. Infants selectively encode the goal object of an actor's reach. *Cognition* 1998;69:1–34. [PubMed: 9871370]
- Woodward AL, Sommerville JA. Twelve-month-old infants interpret action in context. *Psychological Science* 2000;11:73–77. [PubMed: 11228848]

TABLE 1

Behavioural parameters for action events

<i>Event</i>	<i>Action onset</i>	<i>Action completion</i>
Soda	After the bottle has been grasped, the moment the liquid in the bottle starts to move and is tilted towards the glass	The moment the bottle is tilted over the glass and liquid has reached the bottle's spout
Ice cream scoop 1	The scoop is out of the container and the actor's shirt is visible between the bottom of the scoop and ice cream carton	The ice cream scooper is above the glass and the scoop of ice cream begins to slide out from the scoop
Ice cream scoop 2	The scoop is out of the container and the actor's shirt is visible between the bottom of the scoop and the ice cream carton	The ice cream scooper is above the glass and the scoop of ice cream begins to slide out from the scoop
Sprinkles	The sprinkles container is lifted off the table	The sprinkles container is above the glass and sprinkles have reached the container's spout
Cherry	The actor's fingers clear the lip of the jar	The cherry is positioned over the middle of the glass and the actor's index finger is extended for release