HHS Public Access
Author manuscript
J Exp Child Psychol. Author manuscript; available in PMC 2020 November 17.

Published in final edited form as:
J Exp Child Psychol. 2019 March ; 179: 324–336. doi:10.1016/j.jecp.2018.11.020.

Sustained visual attention is more than seeing

Lei Yuan[a,*], **Tian Linger Xu**[b], **Chen Yu**[a,c], **Linda B. Smith**[a,c,d]

[a]Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN 47405, USA

[b]Department of Computer Science and Cognitive Science Program, Indiana University Bloomington, Bloomington, IN 47405, USA

[c]Cognitive Science Program, Indiana University Bloomington, Bloomington, IN 47405, USA

[d]School of Psychology, University of East Anglia, Norwich, Norfolk, UK

Abstract

Sustained visual attention is a well-studied cognitive capacity that is relevant to many developmental outcomes. The development of visual attention is often construed as an increased capacity to exert top-down internal control. We demonstrate that sustained visual attention, measured in terms of momentary eye gaze, emerges from and is tightly tied to sensory-motor coordination. Specifically, we examined whether and how changes in manual behavior alter toddlers' eye gaze during toy play. We manipulated manual behavior by giving one group of children heavy toys that were hard to pick up and giving another group of children perceptually identical toys that were lighter and easy to pick up and hold. We found a tight temporal coupling of visual attention with the duration of manual activities on the objects, a relation that cannot be explained by interest alone. Toddlers in the heavy-object condition looked at objects as much as toddlers in the light-object condition but did so through many brief glances, whereas looks to the same objects were longer and sustained in the light-object condition. We explain the results based on the mechanism of hand–eye coordination and discuss its implications for the development of visual attention.

Keywords

Sustained visual attention; Hand–eye coordination; Multimodal; Perception–action; Manual behavior; Developmental systems

Introduction

The visual world presents a flux of concurrent streams of sensory stimulation. Making sense of all this information requires selecting and sustaining attention on just some of it. Not surprisingly, infants' and children's ability to select and sustain attention on a target is

*Corresponding author.: leiyuan@indiana.edu (L. Yuan).

Appendix A. Supplementary material
Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2018.11.020.

predictive of learning in many domains (Fisher, Godwin, & Seltman, 2014; Kannass & Oakes, 2008; Ruff & Lawson, 1990; Yu & Smith, 2014). In the adult literature, *visual* attention is often studied as a visual process influenced from below by visual properties and from above or top-down by conceptual knowledge (Buschman & Miller, 2007; Egeth & Yantis, 1997; Posner, 1980). However, looking, or directing gaze to a target, is also an action that is tightly coordinated with other actions (Gibson, 1963, 2015). Many developmentalists have argued that manual actions provide a context for sustaining—and learning to sustain—visual attention on objects (Needham, Barrett, & Peterman, 2002; Ruff, 1989; Yu & Smith, 2012, 2016b). Considerable evidence indicates that manual actions select information for attention and that manual actions are also associated with more enduring attention to an object (Hayhoe & Ballard, 2005; Yu, Smith, Shen, Pereira, & Smith, 2009). The central question for the current study was the mechanism through which manual actions provide the context for the development of sustained attention.

The starting point for the current study and the tested hypotheses is the seminal work of Ruff (1986; see also Ruff & Lawson, 1990). Her program of research defined sustained attention in terms of not just continuous visual attention to the object but also including toddlers' holding of the attended object. Sustained attention, defined in these terms of hand and eye measures, has been shown to increase incrementally from toddlerhood through the preschool period and to predict future attention, self-regulation, and vocabulary development (Kannass & Oakes, 2008; Lawson & Ruff, 2004; Razza, Martin, & Brooks-Gunn, 2010). More recent research using head-mounted eye trackers has confirmed the link between toddlers' hand actions and gaze (Pereira, Smith, & Yu, 2014; Yu & Smith, 2014, 2016a). Toddlers' visual attention to an object is more enduring, with positive consequences of a better memory for the object and associated name, when the visually attended object was also in contact with the hands (Pereira et al., 2014). Other studies indicate a tight temporal relation between hands and eyes; when engaged in active play, toddlers' hands and eyes are dynamically coupled and move nearly simultaneously to the same object (Yu & Smith, 2013, 2017). This fact suggests the hypothesis tested in this study—namely that sustained gaze on an object may emerge early in development as a multimodal event and may be maintained by the *joint* sensory-motor inputs provided by eyes and hands. The hypothesis was not that hands temporally lead eyes to an object (given that eyes generally, but not always, lead hands to the object; Pelz, Hayhoe, & Loeber, 2001) but rather that sustained hand actions directly sustain gaze.

Although sustained hand actions could lead to sustained gaze to the object through top-down conceptual effects or goals, the current hypothesis was that the effects of hand actions on gaze arise within the sensory-motor system itself through the real-time dynamic coordination of eye and hand movements. Although one can look at an object without touching it, making manual contact with an object typically requires looking to the object that is to be touched, and continued manual engagement with an object might be expected at the sensory-motor level and might entrain sustained looking. If gaze is tightly coordinated with goal-directed hand actions in this way for toddlers, then toddlers' looking and acting on objects *should be tightly aligned* in time, such that altering the temporal structure of one should lead to corresponding changes of the other. This sensory-motor hypothesis fits evidence from studies of visual attention in human adults and primates, which shows direct effects of bodily actions on both gaze and the internal processes that underlie visual attention (Thura, Boussaoud, & Meunier, 2008). For example, eye movements (Grosbras, Laird, & Paus, 2005), head movements (Colby & Goldberg, 1999), and hand movements (Thura, Hadj-Bouziane, Meunier, & Boussaoud, 2008) bias visual attention in the direction of the movement.

The alternative hypothesis is that the association between handling objects and sustained attention in toddlers derives not from direct sensory-motor coupling but rather from top-down motivation and conceptual factors. For example, an object that is interesting to look at is likely to also be interesting to touch and hold. If this is the case, then the overall duration of interest in an object, as measured by gaze and hand contact, may be expected to be correlated, but there is no strong prediction of direct dynamic coordination in time—that gaze and hands should move to and away from the object together in time. More specifically, by the sensory-motor hypothesis, if we alter the dynamics of hand contact to the object, then we should alter gaze dynamics as well. By the alternative hypothesis that hand engagement and eye engagement are driven separately by top-down goals, altering the dynamics of manual engagement need not alter the dynamics of gaze.

We tested these predictions by creating two identical sets of toys—designed and shown through pilot testing—to have properties that encourage manual exploration through a variety of hand actions. For both groups of children, the objects had holes and moveable parts that invited manual actions. For one group of children, the objects were light and easy to hold while being acted on. These toys should elicit long-lasting bouts of manual contact. For the other group of children, the same objects had weights put in them so that they were heavy and hard to hold. For these objects, we expected hand actions to emerge predominantly as a series of touches and pokes as the heavy objects sat on the table. Fig. 1 illustrates the main hypothesis and the alternative possibility. Fig. 1A shows a stream of hand events in time, with each rectangle representing unbroken hand contact with an object. In the light-object condition, because toddlers could hold the object while manually exploring it, the expectation was that the duration of hand contact would be relatively long. In the heavy-object condition, because toddlers were expected to have difficulty in holding the object, the expectation was that hand contact would occur in a series of brief touches, pokes, and handling. Given these two different expected hand activities, the key prediction concerned the dynamics of gaze. If looking at objects and acting on objects are tightly aligned temporally, then altering the temporal structure of hand contact should lead to corresponding changes in the temporal structure of gaze, such that continuous hand contact should support continuous eye contact and bursts of intermittent hand contact with a single object should support bursts of intermittent eye contact to that object—the temporal alignment hypothesis. Thus, as illustrated in Fig. 1B, unbroken look durations should be longer in the light-object condition than in the heavy-object condition. The alternative possibility of shared top-down goals that independently affect eye and hand actions— what we call the “interest” hypothesis—is illustrated in Fig. 1C. When children are manually engaged with a single object, regardless of whether that engagement consists of continuous hand contact or bursts of hand activity with the same object, gaze will stay focused on the object of interest.

Method

Participants

The final sample consisted of 31 parent–toddler dyads (toddler mean age = 21 months, range = 18–25). Roughly half ($n = 16$) of the toddlers were assigned to play with lightweight toys, whereas the others ($n = 15$) played with heavyweight toys. Data from 1 child who completed the study were not included because of equipment failure. Children were recruited from a population of working- and middle-class families in a U.S. midwestern town. All parents provided informed consent. The sample size was adequate due to the high-density nature of the data—all participants wore a head-mounted eye tracker that sampled at 30 Hz and contributed 11,368 frames of data on average—and based on prior work using a similar data collection method (Yu & Smith, 2013, 2016a).

Stimuli

Fig. 2A shows the two sets of six novel toys that were developed from extensive pilot work to be engaging for manual play with moveable elements, openings, and possible actions. They were made of hardened clay, painted red, blue, or green, and were roughly the same size (9.5 × 6.5 × 5 cm). The two sets were identical in terms of shape, size, and color, with the only difference being their weight. The heavy toys (constructed by putting weights inside) averaged 639 g, seven times heavier than the average weight of the toys in the light set, which was 91 g.

Context

Because toddlers do not play with toys as consistently or happily when not with their parents, toy play was with one parent. As shown in Fig. 2B, the parent and child sat across a small table (61 × 91 × 64 cm). The child sat in a small chair, and the parent sat cross-legged on a pillow. The infant (and parent) wore a head-mounted eye tracker with a sampling rate of 30 Hz (Positive Science, Rochester, NY, USA) (see also Franchak, Kretch, Soska, & Adolph, 2011). The eye tracker consisted of a scene camera that captured the egocentric view of the participant and an infrared camera that was mounted on the head, pointed to the right eye of the participant, and recorded the eye-in-head position (x and y) in the captured scene. Another high-resolution camera (recording rate of 30 Hz) was mounted above the table and provided a bird's-eye view that was independent of the participant's movements.

Procedures

To place the eye tracker on a child's head, one experimenter attracted the child's attention with an interesting toy while another experimenter put the eye-tracking gear low on the child's forehead. To calibrate the eye tracker, the experimenter directed the child's eyes toward an interesting toy, which was repeated 15 times while the toy was placed at various locations on the table. Parents were told that the goal of the experiment was to study how parents and their toddlers naturally interact during toy play and were instructed to play with their toddlers as naturally as possible. The free-play session lasted for a total of 6 min and was composed of four trials (each lasted 1.5 min). The six novel toys were grouped into two sets (A and B). Each set had three different colored objects (red, blue, and green). The sets were interleaved, and the order was counterbalanced across dyads. At the end of each trial, the experimenter signaled the parent with a clicking sound and quickly replaced the old set of toys with a new set.

Coding

The eye tracker sampled at 30 frames per second during the 6-min play session, yielding a theoretical 10,800 frames of data. There was no significant difference in the number of recorded frames between the two experimental conditions, $t(30) = 0.51$, $p = .61$, and the total final sample included a corpus of 352,417 frames of data. Three regions of interest (ROIs) were defined for the gaze data: the green, blue, and red objects. These ROIs were coded manually by naïve coders who annotated frame by frame when the participant fixated at any of the three ROIs. Another coder independently coded 29% of the frames, and there was 86% agreement between coders (Cohen's kappa = .81). Hand contact was coded based on the frame-by-frame images captured by the eye tracker. Three ROIs were defined as when the participant was in manual contact with any of the three objects: the green, blue, or red object. The coders also consulted the third-person view camera in the case of uncertainty (e.g., the physical contact between a hand and an object could not be reliably determined). Another coder independently coded 29% of the frames, and there was 98% agreement between coders (Cohen's kappa = .96). These resulted in two temporally synced streams of data: eye gaze and manual action. Fig. 2C provides sample data from 1 participant (top row: child's eye gaze on the objects; middle row: child's left hand in contact with the objects; bottom row: child's right hand in contact with the objects). In all following analyses, data from right and left hands were coded individually and then combined as manual contact defined as either hand or both hands. The duration of each event or bout was defined as a *continuous* hand contact or gaze on an object and was calculated by summing all frames within the event.

All of the analyses reported in the Results section focused on children's behavior only as it is directly related to our hypotheses. However, because social partners interact during play, and children's behavior could be affected by parents' behavior and not the specific manipulation of the toys, we also coded and analyzed parents' gaze and hand contact data. We found no significant difference in parents' behavior between the two conditions. These results are included in the online supplementary material.

Results

We first report on infants' hand contact with the objects and then turn to the main question—whether the dynamics of gaze are aligned with the dynamics of hand actions.

The dynamics of manual actions

As shown in Table 1, the hands of children in the heavy-object and light-object conditions contacted the objects for comparable amounts of total time, $t(29) = 0.36$, $p = .70$, and there was no significant difference in the proportions of total play time that children in the light-object and heavy-object conditions were in hand contact with the objects, $t(29) = 0.24$, $p = .80$. Furthermore, hand contact bouts from both conditions were distributed across the three toys. In terms of total amount of hand contact, then, the toys in both the heavy-object and light-object conditions were comparably engaging to the children. However, as expected, the dynamics of hand contact differed across the two conditions. Fig. 3A (left) shows the frequency (counts per minute) distribution of the durations of hand contact in both conditions. As with most natural behaviors, these distributions were not normal but rather skewed, such that most hand contacts were very brief, but some were quite long. Statistical analyses assuming normal distributions of data (e.g., *t* test) are not appropriate for these extremely skewed distributions (Gibbons & Chakraborti, 2011) because there is no central tendency. Accordingly, our analysis plan for the durations of hand contact and subsequently for the durations of looking used two approaches. The first is an event-level analysis in which we compared the entire frequency distribution of all object contact durations from the two conditions collapsed across children. As can be seen in Fig. 3A (left), a Mann–Whitney–Wilcoxon test revealed that the distributions of durations in the two conditions were significantly different (U = 1,800,000, $p < .001$); as expected, the durations of manual actions were briefer in the heavy-object condition (*Mdn* = 0.80 s, M = 2.43 s) than those in the light-object condition (*Mdn* = 1.16 s, M = 3.29 s).

The second analysis was an individual-level analysis in which we constructed two scores for each child: the proportion of all looks that fell at two duration categories: very brief durations (the head of the right-skewed distributions; <1 s) and substantially longer durations (the tail of the distribution; >3 s, the threshold for sustained visual attention used by Ruff & Lawson, 1990, and by Yu & Smith, 2016a, 2016b). Although we leave out the middle durations, these two measures (accounting for 78% of data and normalized by participant for the total number of acts) are dependent. Therefore, we also report a secondary analysis based on the total number of events rather than a proportion. A linear mixed-effects regression model was conducted using the *lme*4 package in R (Version 3.0.1; Bates, Mächler, Bolker, & Walker, 2015); the p values for regression coefficients were obtained using the car package (Fox & Weinberg, 2011). Condition (heavy-object vs. light-object) and event type (brief vs. sustained) were submitted as fixed effects, and participant was submitted as a random effect. As shown in Fig. 3A (right), there was a significant interaction between condition and event type ($b = -0.05$, $SE = 0.02$, $p = .037$). The heavy-object condition had a higher proportion of brief looking events than the light-object condition (52% vs. 46%), but the light-object condition had a higher proportion of sustained looking events than the heavy-object condition (30% vs. 25%). The same analysis, when conducted on the total number of hand contacts in each duration category, yielded the same significant interaction between condition and event type ($b = -11.69$, $SE = 5.63$, $p = .037$). In sum, children in the light-object condition picked up and held objects, resulting in fewer but longer manual contact events; children in the heavy-object condition generated more but briefer manual contacts with the objects.

Fig. 4 (top panel) shows our expectation of how touches would be distributed in the heavy-object condition—repeated brief touches and pokes to the same object. However, it is also possible that the heaviness of the objects could disrupt play more dramatically, such that children in the heavy-object condition often switched between objects (e.g., a touch to the blue object followed by a touch to the red object). The evidence supports our expected pattern. There was no significant difference in the frequency of switches between different objects in the heavy-object condition (*M* = 5.75, *SD* = 2.27) and the light-object condition (*M* = 5.12, *SD* = 2.4), *t*(29) = 0.74, *p* = .46. However, the frequency of repeated hand contact to same object was significantly higher in the heavy-object condition (*M* = 16.01, SD = 6.67) than in the light-object condition (*M* = 10.51, *SD* = 3.76), *t*(29) = 2.85, *p* = .008, *d* = 1.01, with hand activity in the heavy-object condition consisting of a series of touches to the same object. Thus, toddlers in the heavy-object condition often touched the same object repeatedly in short bursts, whereas toddlers in the light-object condition often maintained hand contact with the explored object for a long time.

These results set the stage for the main question: Given that hand dynamics differ between the two conditions, do eye dynamics differ as well? Is gaze to objects in shorter bursts when the objects are heavy but sustained when the objects are lighter and, thus, in longer contact with hands?

The dynamics of gaze

As shown in Table 1, children in the light-object and heavy-object conditions looked at the objects for comparable total amounts of time, *t*(29) = 0.24, *p* = .80, and had comparable proportions of play time spent looking to the objects, *t*(29) = 0.52, *p* = .60. Furthermore, looking events from both conditions were distributed across the three toys. As predicted, the dynamics of gaze differed between the two conditions. Fig. 3B (left) shows the frequency (counts per minute) distribution of the durations of all individual looking events (across all children) for each condition. A Mann–Whitney–Wilcoxon test revealed a significant difference between the distributions (*U* = 1,900,000, *p* < .001), supporting the hypothesis; looking events were briefer in the heavy-object condition (*Mdn* = 1.03 s. *M* = 1.90 s) than in the light-object condition (*Mdn* = 1.26 s, *M* = 2.30 s). For the individual-level analysis, we constructed two scores for each child: the proportion of all looks that fell into two duration categories: very brief durations (the head of the right-skewed distributions; <1 s) and substantially longer durations (the tail of the distribution; >3 s, a common threshold for sustained attention used by Ruff & Lawson, 1990, and Yu & Smith, 2016a, 2016b). A linear mixed-effects regression model was conducted in which condition (heavy-object vs. light-object) and event type (brief vs. sustained) were submitted as fixed effects and participant was submitted as a random effect. As shown in Fig. 3B (right), there was a significant interaction between condition and event type (*b* = −0.07, *SE* = 0.02, *p* < .001). The heavy-object condition had a higher proportion of brief looking events than the light-object condition (48% vs. 41%), but the light-object condition had a higher proportion of sustained looking events than the heavy condition (25% vs. 20%). The same analysis, when conducted on the total number of looking events in each duration category, yielded the same significant interaction (*b* = 10.60, *SE* = 4.34, *p* = .02).

In sum, the results of the looking patterns mirror the results of the hand contact activity; children in the heavy-object condition produced more rapid but frequent hand contact events, as well as more rapid but frequent looking events, compared with children in the light-object condition. Thus, sustained hand contact is associated with sustained gaze. In other words, by the definitions of sustained attention (continuous gaze to an object longer than 3 s) used in previous research (Ruff & Lawson, 1990; Yu & Smith, 2016b), children in the light-object condition showed more sustained attention than children in the heavy-object condition.

The dynamics of gaze during manual action

By hypothesis, the common dynamics of hand and eye in the two conditions reflects a direct effect, namely that hand contact sustains eye contact because the two actions are temporally coordinated within the sensory-motor system. Thus, long durations of hand contact in the light-object condition should coincide with long durations of gaze to *the same object*, and short durations of hand contact in the heavy-object condition should *coincide* with short durations of gaze *to the same object*, a key component of the predictions not addressed in the above analyses. To test this prediction, we measured the durations of joint hand and eye events directed to the same object. Fig. 3C (left) shows the frequency distributions of looking events during hand contact, and a Mann–Whitney–Wilcoxon test indicated that the distributions were significantly different between the two conditions (U = 1,400,000, $p < .001$); joint eye–hand events were briefer in the heavy-object condition (*Mdn* = 0.50 s, M = 1.00 s) than in the light-object condition (*Mdn* = 0.70 s, M = 1.30 s). For each child, we further calculated the proportion of those joint eye–hand events that were less than 1 s and those that were more than 3 s. A linear mixed-effects regression model was conducted in which condition (heavy-object vs. light-object) and event type (brief vs. sustained) were submitted as fixed effects and participant was submitted as a random effect. As shown in Fig. 3C (right), there was a significant interaction between condition and event type (b = −0.06, SE = 0.01, $p < .001$); the heavy-object condition had a higher proportion of brief joint eye–hand events than the light-object condition (67% vs. 59%), but the light-object condition had a higher proportion of sustained joint eye–hand events than the heavy-object condition (12% vs. 8%). In sum, the heavy-object condition is characterized by shorter joint hand–eye events in which toddlers' hands and gaze were on the object at the same time, and the light-object condition is characterized by longer joint hand–eye events in which toddlers' hands and gaze were on the object at the same time.

Discussion

The main finding is this: Altering the temporal dynamics of manual action led to corresponding changes in the temporal dynamics of visual attention. Past research (Ruff & Lawson, 1990; Ruff, 1986, 1989) indicated an association between toddlers' holding of an object and sustained visual attention, with holding interpreted as a sign of effortful focused attention. In the current study, children who played with light and easy to hold toys showed both continuous hand contact and more sustained visual attention, behavior fitting the prior characterization of focused attention. However, children who played with just as interesting but heavy toys manually engaged with those toys in briefer bursts of hand activity and, because gaze was coupled to their hand actions, also showed less sustained attention. These facts suggest that continuous hand contact might not be a mere sign of focused attention but instead may play an instrumental role in sustaining the duration of gaze to an object. These results also suggest that sustained attention is more likely with objects that support prolonged manual contact than with those that do not. If object play provides a context not just for measuring sustained attention but also for its development (Wass, Porayska-Pomsta,

& Johnson, 2011; Yu & Smith, 2016b), then these objects that support longer manual contact may be a key training ground for the development of sustained attention.

The findings also raise a new mechanistic route through which toddlers may control their visual attention through their own actions. Visual attention is typically thought of as determined exogenously by the attention-getting properties of the visual stimulus or endogenously through top-down control (Colombo, 2001; Emberson, 2017; Richards & Casey, 1992; Ruff & Capozzoli, 2003). But neither of these routes seems to explain the current results. Children in the heavy-object condition, who played with the objects just as much as children in the light-object condition, could have sustained their gaze on the acted-on object while their fingers and hands poked and jabbed the object during their manual explorations, but they did not. Instead, the dynamics of hands and eyes were aligned in more rapid bursts to the objects, whereas the aligned dynamics of the hands and eyes of the children in the light-object condition included more enduring hand contact and gaze. There is no easy explanation of this finding through traditional ideas of exogenous or top-down control given that toddlers visually attended to and manually engaged with the objects for comparable total durations across the play session. Instead, the current findings were predicted from and implicate a multimodal pathway in which the coactivation of a second sensory-motor system directed to the same object—hand contact—entrains gaze so that longer hand contact sustains longer looks and shorter hand contact is associated with shorter looks to the object.

Looking—directing and maintaining gaze to an object—is a motor act. A large body of literature has documented the role of engaged manual and visual exploration of objects in supporting stabilized and aligned heads, eyes, and posture (e.g., Bertenthal & von Hofsten, 1998; Saavedra, Woollacott, & van Donkelaar, 2010; Soska, Galeon, & Adolph, 2012), with positive consequences for visual learning about objects (Baumgartner & Oakes, 2013; Soska, Adolph, & Johnson, 2010; Woods & Wilcox, 2013), visual attention (Needham et al., 2002), and joint attention (de Barbaro, Johnson, Forster, & Deák, 2016; Yu & Smith, 2014). Other evidence shows that holding objects stabilizes posture and head movements in early sitters and walkers (Claxton, Melzer, Ryu, & Haddad, 2012; Claxton, Strasser, Leung, Ryu, & O'Brien, 2014). We propose that through perhaps similar intersensory processes, holding objects stabilizes the motor act of gaze to an object. This hypothesis of a direct multisensory pathway in sustaining attention fits with evidence from adults showing considerable manual–visual interactions in behavior and neural processing (Abrams, Davoli, Du, Knapp, & Paull, 2008; Macaluso & Driver, 2001; Macaluso, Frith, & Driver, 2000; Park & Reed, 2015; Taylor-Clarke, Kennett, & Haggard, 2004) and developmental findings and theories on the especially important role of intersensory interactions in the early developmental process (Bahrick & Lickliter, 2000; Brenna et al., 2015; Lewkowicz & Lickliter, 1994).

Because sustained attention during late infancy is predictive of future development, and because deficits in sustained attention are markers for later diagnoses of attentional disorders (Barkley, 1997; Jones, Rothbart, & Posner, 2003), the current findings have direct consequences for understanding the developmental origins of individual differences in sustained attention (Iverson, 2010). There are growing suggestions that the development of sustained visual attention during infancy has a strong experiential component, and individual differences may arise from experiences that depend on sensory-motor coordinations, including those between hands and eyes (Wass et al., 2011; Yu & Smith, 2014). This idea is consistent with findings showing that sensory-motor dis-coordinations are often the earliest signs of atypical development and predictive of long-term outcomes across many domains (D'Souza, Cowie, Karmiloff-Smith, & Bremner, 2017; Provost, Lopez, & Heimerl, 2007). It is also consistent with correlational and experimental studies linking object manipulation to object name learning, object memory, and object attention (James, Jones, Smith, & Swain, 2014; Needham et al., 2002). The current findings suggest the potential value of objects and tasks that invite and sustain manual contact as a possible malleable factor in supporting sustained attention.

In conclusion, the current findings illustrate how visual attention develops in a larger network of behaviors that involve much more than vision itself (Byrge, Sporns, & Smith, 2014). The tight coordination of hands and eyes in toddlers has direct real-time effects on sustained visual attention, and we propose that these may support the development of visual attentional skills. The larger idea that development in one domain depends on and supports developments in other domains—even ones seemingly far away—is consistent with developmental system views (Gottlieb, 2007; Oyama, Griffiths, & Gray, 2003; Thelen & Smith, 1994) and the role of behavior in the development of functional and structural brain networks (Byrge et al., 2014; Herholz & Zatorre, 2012). Visual attention in real time emerges from at least two sensory modalities—haptic and visual—and two motor actions—manipulation and gaze—that are dynamically coordinated during moment-by-moment engagement with objects. Unbroken manual contact with an object provides the context for unbroken gaze on the object. The real-time effects of sustained holding and looking over developmental time may build the circuitry that supports internally driven sustained visual attention to a target. This is a key hypothesis for future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by National Institutes of Health Grants R01 HD074601 awarded to CY, R21 EY017843 awarded to LBS and F32 HD090827-02 awarded to LY. We thank Melissa Elston, Charlotte Wozniak, Melissa Hall, Charlene Tay, Steven Elmlinger, Madison Philips, Anting Chen, and Seth Foster for data collection.

References

Abrams RA, Davoli CC, Du F, Knapp WH, & Paull D (2008). Altered vision near the hands. Cognition, 107, 1035–1047. [PubMed: 17977524]

Bahrick LE, & Lickliter R (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. Developmental Psychology, 36, 190–201. [PubMed: 10749076]

Barkley RA (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. Psychological Bulletin, 121, 65–94. [PubMed: 9000892]

Bates D, Mächler M, Bolker B, & Walker S (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67, 1–48.

Baumgartner HA, & Oakes LM (2013). Investigating the relation between infants' manual activity with objects and their perception of dynamic events. Infancy, 18, 983–1006.

Bertenthal B, & von Hofsten C (1998). Eye, head and trunk control: The foundation for manual development. Neuroscience & Biobehavioral Reviews, 22, 515–520. [PubMed: 9595563]

Brenna V, Nava E, Turati C, Montirosso R, Cavallini A, & Borgatti R (2015). Intersensory redundancy promotes visual rhythm discrimination in visually impaired infants. Infant Behavior and Development, 39, 92–97. [PubMed: 25827259]

Buschman TJ, & Miller EK (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science, 315, 1860–1862. [PubMed: 17395832]

Byrge L, Sporns O, & Smith LB (2014). Developmental process emerges from extended brain–body–behavior networks. Trends in Cognitive Sciences, 18, 395–403. [PubMed: 24862251]

Claxton LJ, Melzer DK, Ryu JH, & Haddad JM (2012). The control of posture in newly standing infants is task dependent. Journal of Experimental Child Psychology, 113, 159–165. [PubMed: 22683016]

Claxton LJ, Strasser JM, Leung EJ, Ryu JH, & O’Brien KM (2014). Sitting infants alter the magnitude and structure of postural sway when performing a manual goal-directed task. Developmental Psychobiology, 56, 1416–1422. [PubMed: 24604626]

Colby CL, & Goldberg ME (1999). Space and attention in parietal cortex. Annual Review of Neuroscience, 22, 319–349.

Colombo J (2001). The development of visual attention in infancy. Annual Review of Psychology, 52, 337–367.

D’Souza H, Cowie D, Karmiloff-Smith A, & Bremner AJ (2017). Specialization of the motor system in infancy: From broad tuning to selectively specialized purposeful actions. Developmental Science, 20, e12409.

de Barbaro K, Johnson CM, Forster D, & Deák GO (2016). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother–infant–object interactions. Child Development, 87, 494–512. [PubMed: 26613383]

Egeth HE, & Yantis S (1997). Visual attention: Control, representation, and time course. Annual Review of Psychology, 48, 269–297.

Emberson LL (2017). How does experience shape early development? Considering the role of top-down mechanisms In Benson JB (Ed.). Advances in child development and behavior (Vol. 52, pp. 1–41). London: Elsevier. [PubMed: 28215282]

Fisher AV, Godwin KE, & Seltman H (2014). Visual environment, attention allocation, and learning in young children: When too much of a good thing may be bad. Psychological Science, 25, 1362–1370. [PubMed: 24855019]

Fox J, & Weinberg S (2011). Multivariate linear models in R [appendix to An R companion to applied regression]. Thousand Oaks, CA: Sage.

Franchak JM, Kretch KS, Soska KC, & Adolph KE (2011). Head-mounted eye tracking: A new method to describe infant looking. Child Development, 82, 1738–1750. [PubMed: 22023310]

Gibbons JD, & Chakraborti S (2011). Nonparametric statistical inference In Lovric M (Ed.), International encyclopedia of statistical science (pp. 977–979). Berlin: Springer.

Gibson EJ (1963). Perceptual learning. Annual Review of Psychology, 14(1), 29–56.

Gibson JJ (2015). The ecological approach to visual perception: Classic edition (Vol. 20) London: Routledge.

Gottlieb G (2007). Probabilistic epigenesis. Developmental Science, 10, 1–11. [PubMed: 17181692]

Grosbras M-H, Laird AR, & Paus T (2005). Cortical regions involved in eye movements, shifts of attention, and gaze perception. Human Brain Mapping, 25, 140–154. [PubMed: 15846814]

Hayhoe M, & Ballard D (2005). Eye movements in natural behavior. Trends in Cognitive Sciences, 9, 188–194. [PubMed: 15808501]

Herholz SC, & Zatorre RJ (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. Neuron, 76, 486–502. [PubMed: 23141061]

Iverson JM (2010). Developing language in a developing body: The relationship between motor development and language development. Journal of Child Language, 37, 229–261. [PubMed: 20096145]

James KH, Jones SS, Smith LB, & Swain SN (2014). Young children's self-generated object views and object recognition. Journal of Cognition and Development, 15, 393–401. [PubMed: 25368545]

Jones LB, Rothbart MK, & Posner MI (2003). Development of executive attention in preschool children. Developmental Science, 6, 498–504.

Kannass KN, & Oakes LM (2008). The development of attention and its relations to language in infancy and toddlerhood. Journal of Cognition and Development, 9, 222–246.

Lawson KR, & Ruff HA (2004). Early attention and negative emotionality predict later cognitive and behavioural function. International Journal of Behavioral Development, 28, 157–165.

Lewkowicz DJ, & Lickliter R (1994). The development of intersensory perception: Comparative perspectives. Hillsdale, NJ: Lawrence Erlbaum.

Macaluso E, & Driver J (2001). Spatial attention and crossmodal interactions between vision and touch. Neuropsychologia, 39, 1304–1316. [PubMed: 11566313]

Macaluso E, Frith CD, & Driver J (2000). Modulation of human visual cortex by crossmodal spatial attention. Science, 289, 1206–1208. [PubMed: 10947990]

Needham A, Barrett T, & Peterman K (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using "sticky mittens" enhances young infants' object exploration skills. Infant Behavior and Development, 25, 279–295.

Oyama S, Griffiths PE, & Gray RD (2003). Cycles of contingency: Developmental systems and evolution. Cambridge, MA: MIT Press.

Park GD, & Reed CL (2015). Haptic over visual information in the distribution of visual attention after tool-use in near and far space. Experimental Brain Research, 233, 2977–2988. [PubMed: 26126805]

Pelz J, Hayhoe M, & Loeber R (2001). The coordination of eye, head, and hand movements in a natural task. Experimental Brain Research, 139, 266–277. [PubMed: 11545465]

Pereira AF, Smith LB, & Yu C (2014). A bottom-up view of toddler word learning. Psychonomic Bulletin & Review, 21, 178–185. [PubMed: 23813190]

Posner MI (1980). Orienting of attention. The Quarterly Journal of Experimental Psychology, 32, 3–25. [PubMed: 7367577]

Provost B, Lopez BR, & Heimerl S (2007). A comparison of motor delays in young children: Autism spectrum disorder, developmental delay, and developmental concerns. Journal of Autism and Developmental Disorders, 37, 321–328. [PubMed: 16868847]

Razza RA, Martin A, & Brooks-Gunn J (2010). Associations among family environment, sustained attention, and school readiness for low-income children. Developmental Psychology, 46, 1528–1542. [PubMed: 20677860]

Richards JE, & Casey BJ (1992). Development of sustained visual attention in the human infant In Campbell BA, Hayne H, & Richardson R (Eds.), Attention and information processing in infants and adults: Perspectives from human and animal research (pp. 30–60). Hillsdale, NJ: Lawrence Erlbaum.

Ruff HA (1986). Components of attention during infants' manipulative exploration. Child Development, 57, 105–114. [PubMed: 3948587]

Ruff HA (1989). The infant's use of visual and haptic information in the perception and recognition of objects. Canadian Journal of Psychology, 43, 302–319. [PubMed: 2486501]

Ruff HA, & Capozzoli MC (2003). Development of attention and distractibility in the first 4 years of life. Developmental Psychology, 39, 877–890. [PubMed: 12952400]

Ruff HA, & Lawson KR (1990). Development of sustained, focused attention in young children during free play. Developmental Psychology, 26, 85–93.

Saavedra S, Woollacott M, & van Donkelaar P (2010). Head stability during quiet sitting in children with cerebral palsy: Effect of vision and trunk support. Experimental Brain Research, 201, 13–23. [PubMed: 19756550]

Soska KC, Adolph KE, & Johnson SP (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. Developmental Psychology, 46, 129–138. [PubMed: 20053012]

Soska KC, Galeon MA, & Adolph KE (2012). On the other hand: Overflow movements of infants' hands and legs during unimanual object exploration. Developmental Psychobiology, 54, 372–382. [PubMed: 22487940]

Taylor-Clarke M, Kennett S, & Haggard P (2004). Persistence of visual–tactile enhancement in humans. Neuroscience Letters, 354, 22–25. [PubMed: 14698473]

Thelen E, & Smith LB (1994). A dynamic systems approach to the development of cognition and action (MIT Press/Bradford Books Series in Cognitive Psychology). Cambridge, MA: MIT Press.

Thura D, Boussaoud D, & Meunier M (2008). Hand position affects saccadic reaction times in monkeys and humans. Journal of Neurophysiology, 99, 2194–2202. [PubMed: 18337364]

Thura D, Hadj-Bouziane F, Meunier M, & Boussaoud D (2008). Hand position modulates saccadic activity in the frontal eye field. Behavioural Brain Research, 186, 148–153. [PubMed: 17881066]

Wass S, Porayska-Pomsta K, & Johnson MH (2011). Training attentional control in infancy. Current Biology, 21, 1543–1547. [PubMed: 21889346]

Woods RJ, & Wilcox T (2013). Posture support improves object individuation in infants. Developmental Psychology, 49, 1413–1424. [PubMed: 23046431]

Yu C, & Smith LB (2012). Embodied attention and word learning by toddlers. Cognition, 125, 244–262. [PubMed: 22878116]

Yu C, & Smith LB (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. [PubMed: 24236151]

Yu C, & Smith LB (2014). Linking joint attention with hand–eye coordination—A sensorimotor approach to understanding child–parent social interaction In CogSci 2015: Annual conference of the cognitive science society (pp. 2763–2768). Oakbrook Terrace, IL: Cognitive Science Society.

Yu C, & Smith LB (2016b). The social origins of sustained attention in one-year-old human infants. Current Biology, 26, 1235–1240. [PubMed: 27133869]

Yu C, & Smith LB (2016a). Multiple sensory-motor pathways lead to coordinated visual attention. Cognitive Science, 41(Suppl 1), 5–31. [PubMed: 27016038]

Yu C, & Smith LB (2017). Hand–eye coordination predicts joint attention. Child Development, 88, 2060–2078. [PubMed: 28186339]

Yu C, Smith LB, Shen H, Pereira AF, & Smith T (2009). Active information selection: Visual attention through the hands. IEEE Transactions on Autonomous Mental Development, 1(2), 141–151. [PubMed: 21031153]

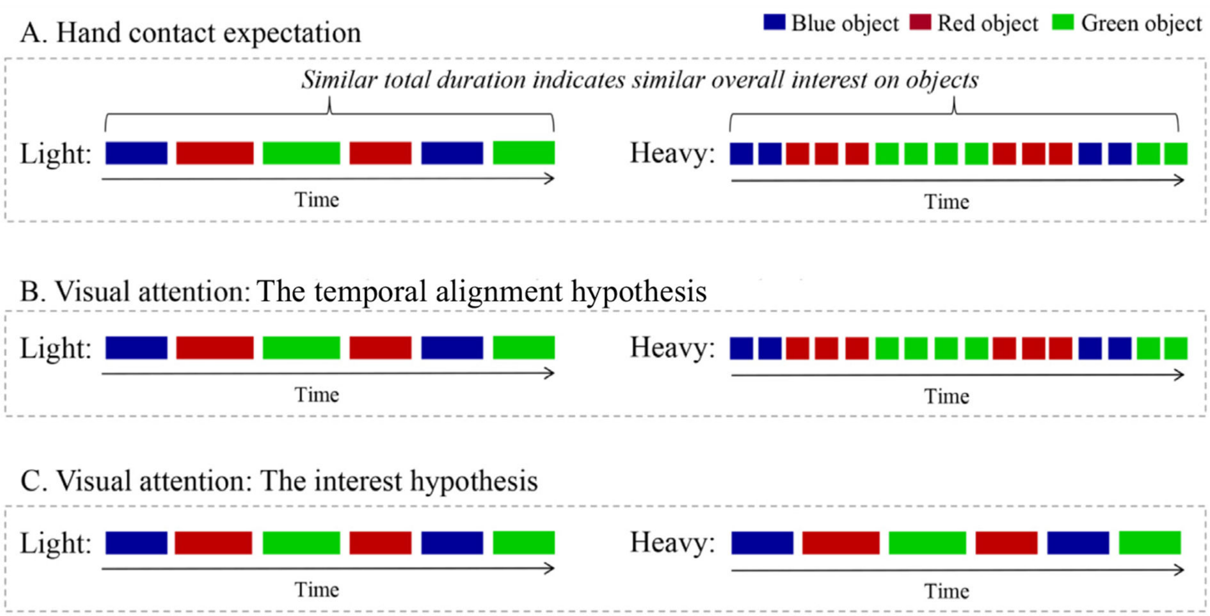

Fig. 1.

Illustrations of the hypotheses. The colored rectangles represent a series of hand and eye contacts with an object; the duration of each contact is indicated by the length of the rectangle. The different colors represent different objects: blue → the blue object; red → the red object; green → the green object. (A) Expectation for hand contact. Children in the heavy-object condition would have more short but frequent hand contacts than those in the light-object condition because of the weight of the objects. (B) Temporal alignment hypothesis for visual attention. If manual actions and gaze are tightly aligned in time, then altering the temporal structure of manual actions would lead to corresponding changes in the temporal structure of gaze. (C) Interest hypothesis for visual attention. If looking behavior is driven by interest, then gaze durations should not differ as long as the hand actions (continuous contact or intermittent contact) remain on the same object.

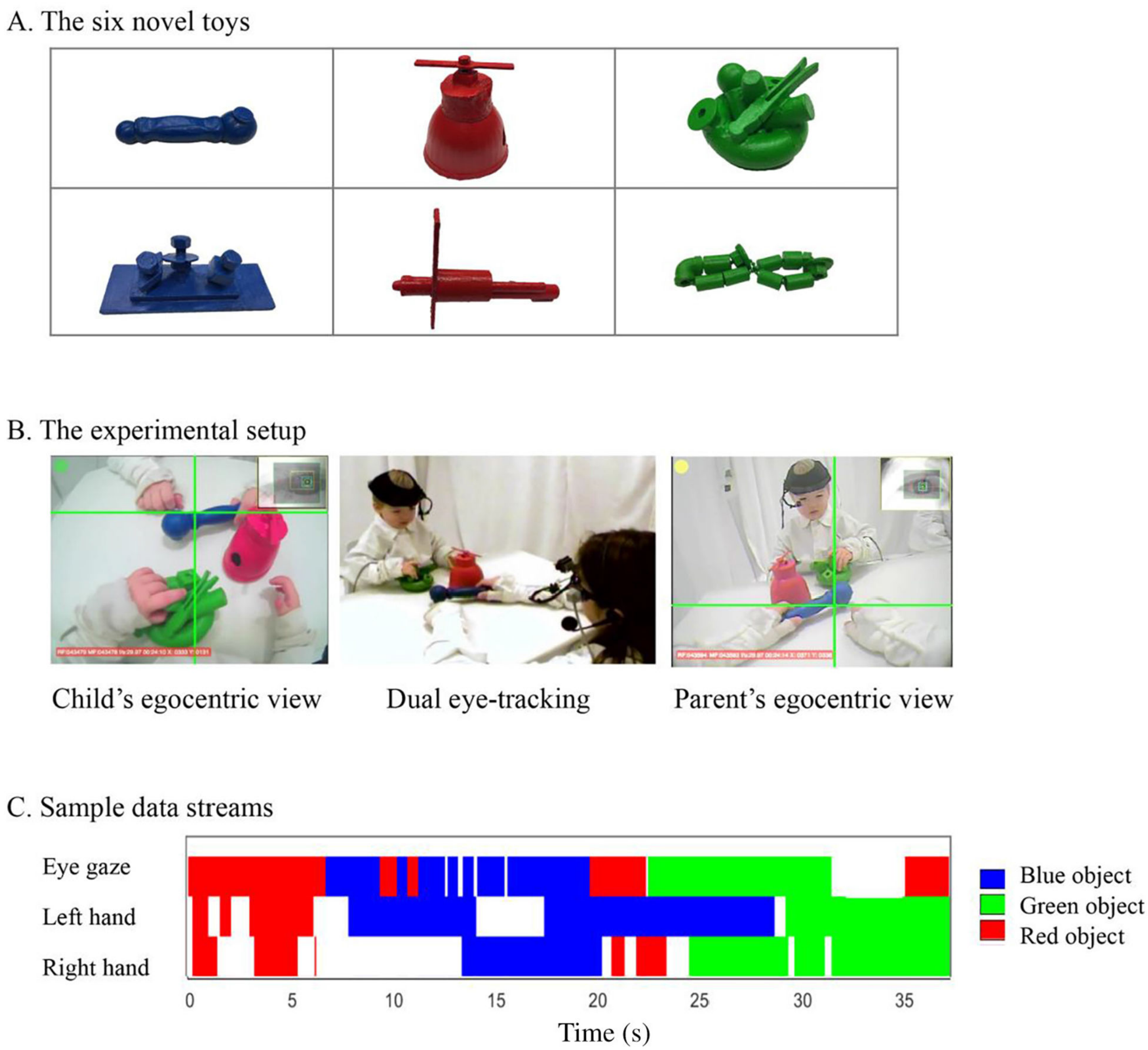

Fig. 2.

Illustrations of the experimental method. (A) The six novel toys used in the study. The heavy and light sets of toys were perceptually identical except for their weight. (B) The experimental setup. Left and right: Eye-tracker images from the child's and parent's egocentric views. Middle: child and parent both wore a head-mounted eye tracker. (C) Sample data from a participant. Three streams of time-locked sensory data—eye gaze (top stream), left-hand hand contact (middle stream), and right-hand hand contact (bottom stream)—are shown from the onset of the experiment to 35 s later. Colors represent the three regions of interest: blue → the blue object; red → the red object; green → the green object.

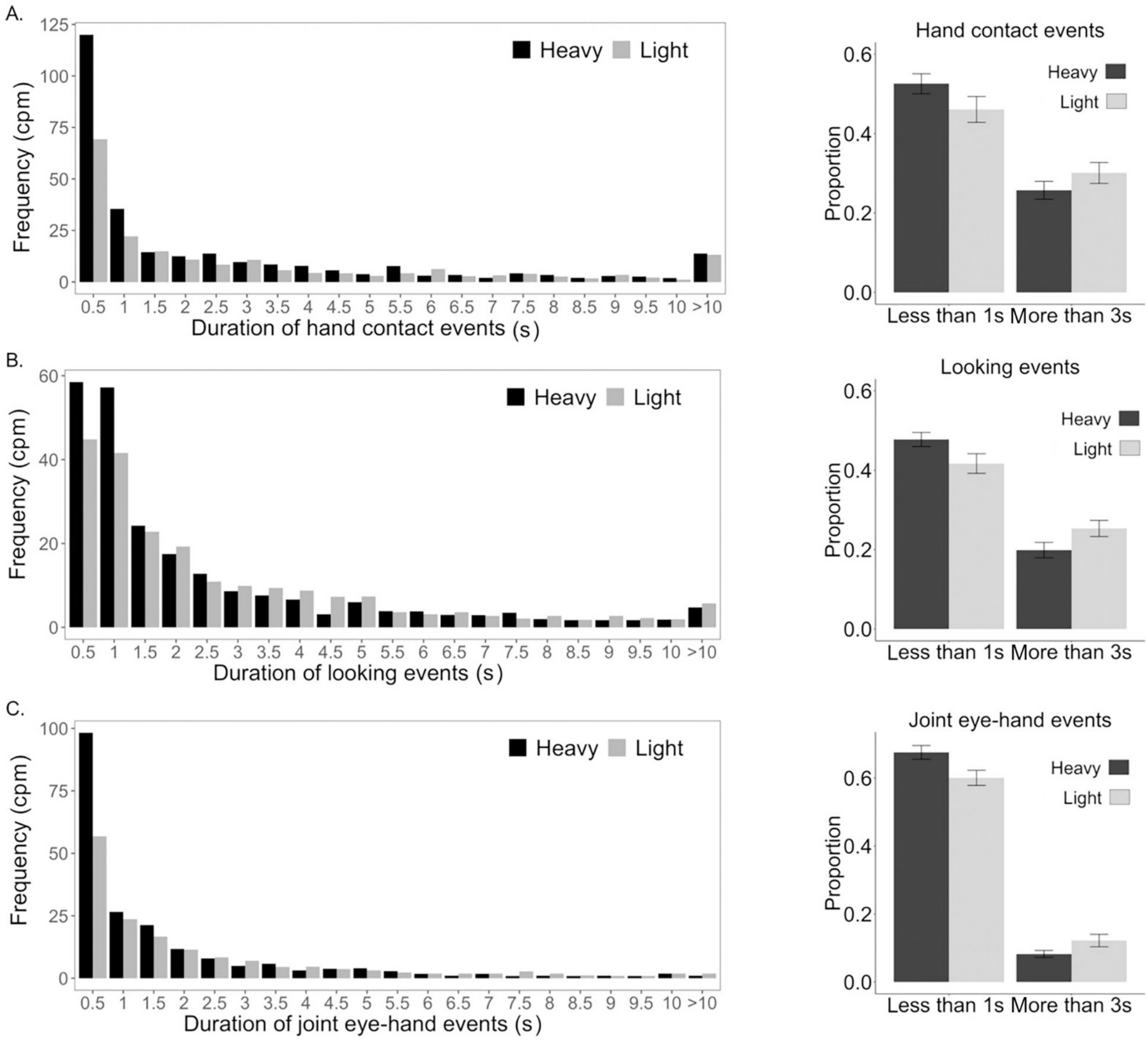

Fig. 3.

Left panels: Frequencies of events that last for different durations. Right panels: Proportions of brief (<1 s) and sustained (>3 s) events in the heavy-object and light-object conditions. Error bars represent standard errors. (A) Hand contact events. (B) Looking events. (C) Joint eye–hand events. cpm, counts per minute.

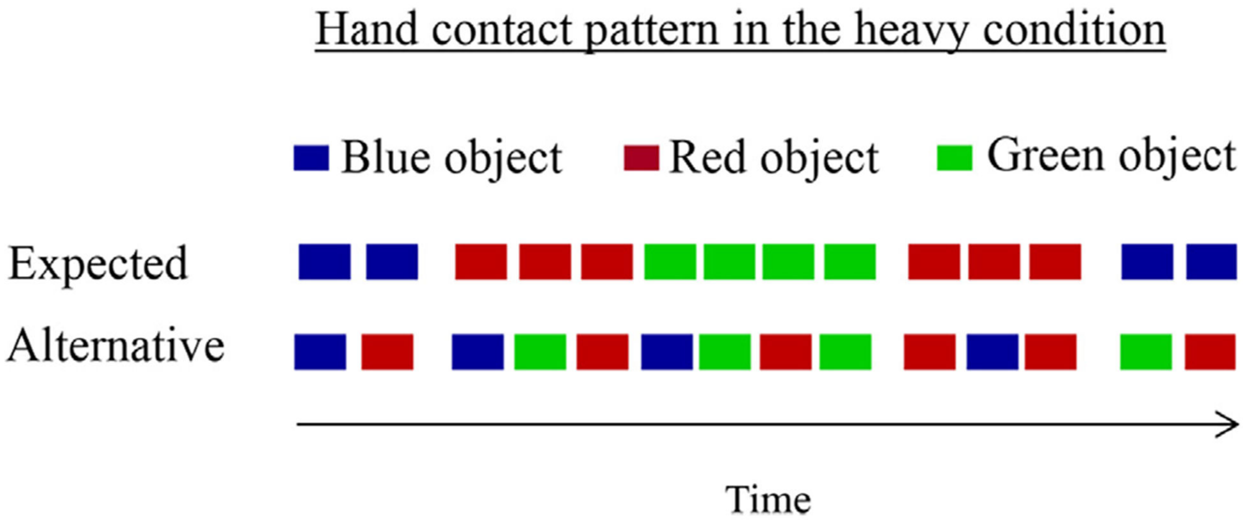

Fig. 4.
Two possible hand contact patterns in the heavy-object condition consistent with briefer durations of hand contact: Top: A series of contacts with one object before switching to another. Bottom: A series of brief contacts on different objects. The pattern in the top row was supported by the empirical data.

Table 1

Overall manual actions and visual attention to objects.

Modality	Condition	Total time (s) (all objects)	Proportion of time (%) (all objects)	Proportion of Object 1 (%)	Proportion of Object 2 (%)	Proportion of Object 3 (%)
Manual actions	Heavy	331(96)	87 (26)	22 (12)	26(10)	40(17)
	Light	348 (117)	85 (30)	31 (18)	29 (16)	25 (8)
Visual attention	Heavy	251 (43)	66 (10)	15(4)	21 (4)	30 (7)
	Light	259 (47)	68 (10)	20 (6)	21 (7)	28 (7)

Note. Standard deviations are in parentheses.