Cognition 177 (2018) 189–197

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Visual short-term memory guides infants' visual attention

Samantha G. Mitsven[a], Lisa M. Cantrell[a], Steven J. Luck[a,b], Lisa M. Oakes[a,b,*]

[a] Center for Mind and Brain, University of California, Davis, United States
[b] Department of Psychology, University of California, Davis, United States

ARTICLE INFO

Keywords:
Visual short-term memory
Infancy
Eye movements

ABSTRACT

Adults' visual attention is guided by the contents of visual short-term memory (VSTM). Here we asked whether 10-month-old infants' ($N = 41$) visual attention is also guided by the information stored in VSTM. In two experiments, we modified the one-shot change detection task (Oakes, Baumgartner, Barrett, Messenger, & Luck, 2013) to create a simplified *cued visual search* task to ask how information stored in VSTM influences where infants look. A single *sample* item (e.g., a colored circle) was presented at fixation for 500 ms, followed by a brief (300 ms) retention interval and then a *test array* consisting of two items, one on each side of fixation. One item in the test array matched the sample stimulus and the other did not. Infants were more likely to look at the non-matching item than at the matching item, demonstrating that the information stored rapidly in VSTM guided subsequent looking behavior.

1. Introduction

Infants' visual environment is complex and cluttered, containing many objects and people—both new and familiar—that come in and out of view. Effectively learning about this complex visual environment involves determining which objects are relevant, directing gaze to those objects, and maintaining gaze on those objects in the face of distractions. Infants' ability to control where and how long they look develops considerably in the first postnatal year (Johnson, 1994). In addition, infants' looking is determined by a number of different factors such as stimulus complexity (Cohen, 1972), novelty (e.g., Welch, 1974; Wetherford & Cohen, 1973), and top-down features such as social content (e.g., Gliga, Elsabbagh, Andravizou, & Johnson, 2009; Gluckman & Johnson, 2013). Other work has shown that *where* infants look is influenced by the physical features of the stimuli such as color (e.g., Dannemiller, 1998) and movement (Volkmann & Dobson, 1976).

Most research on infants' looking has examined their looking behavior in habituation or familiarization procedures, which provide infants with tens of seconds or more to learn about an image and store it in memory. This work has shown that when given adequate time to form a memory representation of the stimulus or stimuli, infants look longer at a previously unseen item (i.e., a novel image) than at the now-familiar item (Oakes & Kovack-Lesh, 2007; Rose, 1981). When infants are given less time to form a representation of the stimuli, they tend to look longer at the familiar item than at a novel one (Hunter, Ross, & Ames, 1982; Rose, 1981). Both patterns of findings demonstrate that information stored in memory over several seconds of learning influences the duration of infants' continued looking at those stimuli versus novel stimuli.

More recently, there has been an interest in moment-to-moment changes in infants' looking. Several studies have examined infants' looking in arrays like those used in studies of visual search in adults. For example, by 6 months infants look first at human faces or other social stimuli rather than at non-social stimuli when presented with arrays of complex, realistic objects (Gliga et al., 2009; Gluckman & Johnson, 2013; Kwon, Setoodehnia, Baek, Luck, & Oakes, 2016). These results show that by the second half of the first year of life, complex information contained in the visual scene can guide infants' eye-movements.

This recent interest in moment-to-moment changes in infants' gaze makes contact with the extensive literature on adults' oculomotor behavior when viewing complex scenes. Several findings are relevant. When viewing complex scenes, adult viewers typically make multiple eye movements per second. Moreover, vision is suppressed during the saccades that separate periods of fixation (Henderson, 2003). As a result, viewers must rapidly create memory representations to span the gaps between fixations (Bays & Husain, 2012). Consider, for example, a viewer looking at a table set out with four cups, four plates, and four spoons. Imagine that the viewer's attention is drawn to a cup, and that the viewer fixates this cup for several hundred milliseconds. If the viewer can use information acquired during this fixation to guide the next several eye movements, he or she can explore the scene in a flexible and systematic manner. In this way, information that is rapidly obtained in each fixation can be used to determine what information

* Corresponding author at: Center for Mind and Brain, 267 Cousteau Pl., University of California, Davis, CA 95618, United States.
E-mail address: lmoakes@ucdavis.edu (L.M. Oakes).

https://doi.org/10.1016/j.cognition.2018.04.016
Received 27 July 2017; Received in revised form 8 April 2018; Accepted 19 April 2018

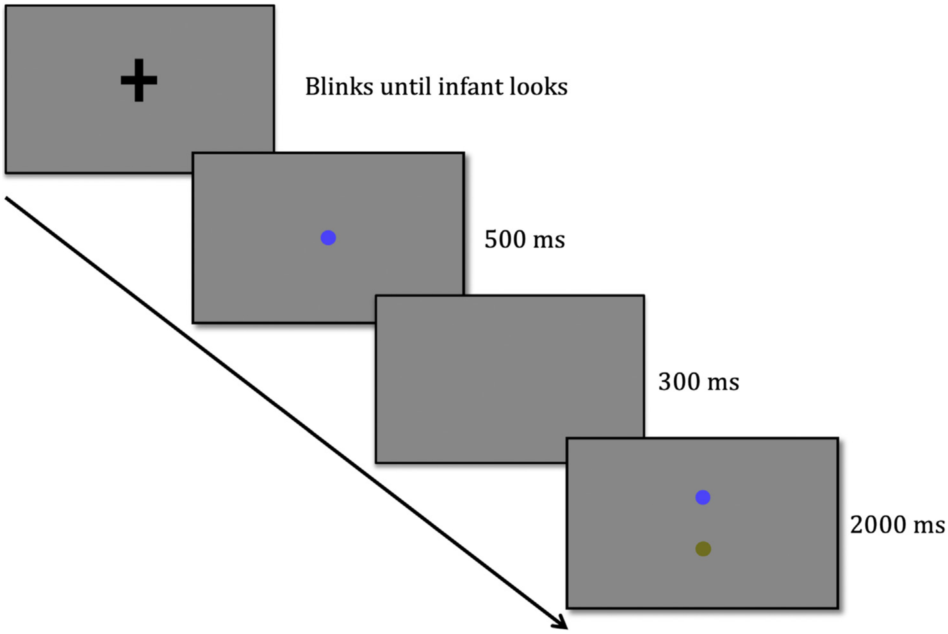

Fig. 1. Schematic illustration of an experimental trial sequence in Experiment 1. In each trial, a sample array of a colored cue circle was presented for 500 ms followed by a brief retention period (300 ms). After the retention period, a visual search array (2000 ms) containing the original cue circle and a circle of a different color was presented with the two items appearing 4.41° from the original cue circle, one on each side of fixation.

should be acquired next.

Many previous studies have shown that adults have this ability: they can rapidly form *visual short-term memory* (VSTM) representations and use them to guide their subsequent shifts of covert and overt attention (Downing, 2000; Downing & Dodds, 2004; Olivers, Meijer, & Theeuwes, 2006; Soto, Hodsoll, Rotshtein, & Humphreys, 2008; Woodman & Luck, 2007). For example, adults can store one or more colors in VSTM at the beginning of a trial and then use these colors to guide eye movements toward items of matching color when performing a visual search task later in the same trial (Beck, Hollingworth, & Luck, 2012; Hollingworth, Matsukura, & Luck, 2013; Soto, Heinke, Humphreys, & Blanco, 2005). Information being held in VSTM can also guide covert shifts of attention in the absence of overt eye movements (Kumar, Soto, & Humphreys, 2009). Importantly, VSTM representations are generated rapidly in adults (e.g., 20–50 ms per item, Vogel, Woodman, & Luck, 2006) and can be used to guide attention within 200 ms (Vickery, King, & Jiang, 2005; Vogel et al., 2006). Thus, adults can rapidly store information in VSTM and then use it to guide their moment-to-moment allocation of both covert and overt attention. This is a key element of the architecture of visual cognition, and it allows the visual system to use the currently fixated information within a scene to guide the further acquisition of information from that scene.

Although significant progress has been made in understanding the ability of infants to rapidly create VSTM representations (Oakes, Baumgartner, Barrett, Messenger, & Luck, 2013; Ross-Sheehy, Oakes, & Luck, 2003), it is not yet known whether infants' cognitive architecture is sufficiently sophisticated to use the perceptual information gained during one brief period of fixation to control oculomotor behavior moments later. This kind of rapid loop between perception and eye movements would be very useful as infants interact with and learn about the visual environment, but it would also require very sophisticated interactions between perceptual and oculomotor systems. Given the results of previous research, it is possible that infants can guide their visual search of a scene only by means of memory representations built up from information acquired over tens of seconds; if true, however, this would severely limit their ability to flexibly explore new scenes. Thus, although we know from previous research that infants can use information stored in memory to guide their eye movements, we do not yet know whether they can take information acquired in one brief period of fixation and use it to guide their allocation of attention moments later. The primary goal of the present investigation was to determine whether infant's cognitive architecture can support the rapid looping between perception and gaze control needed for this ability. We focused on 10-month-old infants, because these infants are clearly old enough to store information robustly in VSTM (Oakes et al., 2013; Ross-Sheehy et al., 2003) and yet young enough that they might not yet have the ability to use this information to rapidly control attention.

Examining this question using controlled laboratory methods will provide insights into how infants acquire information from the complex, dynamic scenes they encounter in their everyday experience. Prioritizing visual input based on its relevance to current goals relies, in part, on the ability to recognize which items in the environment have been previously explored, and are thereby familiar, and which items are novel and require further exploration. Indeed, this is the underlying assumption of *inhibition of return* (IOR, Klein, 2000), or the idea that viewers will inhibit attending to a location that has previously been attended. Even infants show IOR (e.g., Hood & Atkinson, 1993; Valenza, Simion, & Umiltà, 1994), suggesting that moment-to-moment familiarity—at least of an attended location—influences their looking. However, IOR operates on the *locations* that have been attended and not on the information contained at a given location. Consequently, IOR findings do not show that information about the *identity* of the currently fixated location is rapidly stored in VSTM and guides subsequent shifts of attention. In the present work, we ask whether infants' attention is guided in this manner.

To address this question, we created a simplified version of the *cued visual search task* that has been used to study the control of attention by VSTM in adults (Beck et al., 2012; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). The procedure is illustrated in Fig. 1. During an initial sample period, we presented a single item in isolation at fixation (i.e., at the center of the display) for 500 ms, which was followed by a brief (300 ms) retention interval with a blank screen. Finally, we presented a test array of two items (one matching the sample and one different from the sample) at new locations, equidistant from the central fixation point, for 2000 ms (see https://osf.io/5cusz/ for an example of the sequence of events in Experiment 2). The question was whether infants would systematically prefer one of the two items during the test period, indicating that (a) they formed a representation of the initial sample item, and (b) the representation guided their subsequent looking behavior. We required infants to look at the sample item at the beginning of each trial, so they were fixed at the center of the display when the test array appeared and had to make an eye movement away from this