

Multimodal Events and Moving Locations: Eye Movements of Adults and 6-Month-Olds Reveal Dynamic Spatial Indexing

Daniel C. Richardson and Natasha Z. Kirkham
Cornell University

The ability to keep track of locations in a dynamic, multimodal environment is crucial for successful interactions with other people and objects. The authors investigated the existence and flexibility of spatial indexing in adults and 6-month-old infants by adapting an eye-tracking paradigm from D. C. Richardson and M. J. Spivey (2000). Multimodal events were presented in specific locations, and eye movements were measured when the auditory portion of the stimulus was presented without its visual counterpart. Experiment 1 showed that adults spatially index auditory information even when the original associated locations move. Experiments 2 and 3 showed that infants are capable of both binding multimodal events to locations and tracking those locations when they move.

Certain aspects of perception—encoding of object location, binding object properties across modalities, and tracking objects through time and space, for example—have been experimentally isolated and dissected in the literature. By probing the limitations of these systems, researchers have learned much about their possible architectures. A question remains, however, concerning how these abilities operate in concert in the course of everyday life. The brain can simultaneously track the locations of multiple objects (Scholl & Pylyshyn, 1999) and encode in memory thousands of visual images (Standing, Conezio, & Haber, 1970), but to what extent are these operations executed when making a sandwich or

driving to work (Land & Hayhoe, 2001; Land & Lee, 1994)? Aside from the everyday jobs of adult life, the question of how cognitive faculties interact is even more pressing for the task that a developing infant faces in learning about the world.

Simply keeping track of objects can be a challenge. Their absolute location can change, their relative location changes when people move, and their location relative to the eye or hand position changes several times a second. Also, the relevant attributes of objects can be spread across several perceptual modalities, each with its own reference frame. Objects can become occluded, their appearance can change through orientation or lighting differences, and sometimes—through an experimenter's sleight of hand—the features of an object can change completely. With these difficulties, it is less surprising, perhaps, that just keeping track of objects in the task at hand means that people can be blind to many other things around them (Most & Alvarez, 2002; Rensink, 2002). For an infant of 6 months, who is still learning to orientate his or her attention properly (Colombo, 2001), the problem of keeping track of objects must be even more difficult. Yet few abilities could be more important, for if the infant were able to track objects despite these problems, it would assist in learning all sorts of object properties and regularities.

We submit that there is a particular, pragmatic level at which various perceptual abilities cooperate. The present studies investigated how a certain set of faculties work together to produce the behavior that we term *dynamic spatial indexing of multimodal events*. This concept weaves together several strands of the literature. It has been shown that spatial information is encoded with a high degree of automaticity (Nissen, 1985) and that the visual system can dynamically track the locations of multiple objects as they move through space (Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999). *Indexing* refers to the finding that locations are not just encoded but can be used to organize and “look-up” information about objects (Ballard, Hayhoe, Pook, & Rao, 1997) or multimodal events (Richardson & Spivey, 2000). When these strands are entwined, they define dynamic spatial indexing of multimodal events, or simply, spatial indexing. Arguably, it is at this level that this particular cluster of perceptual faculties be-

Daniel C. Richardson and Natasha Z. Kirkham, Department of Psychology, Cornell University.

Authorship is equal, and order of authorship is arbitrary.

A portion of these data were presented as posters at the Levels of Perception Conference, Toronto, Ontario, Canada, June 2001; the Cognitive Development Society Conference, Virginia Beach, Virginia, June 2001; and the International Conference of Infant Studies, Toronto, Ontario, Canada, April 2002. This research was supported by National Science Foundation Grant BCS-0094814 and by National Institute of Child Health and Human Development Grant R01-HD40432 to Scott P. Johnson.

We thank Michael Spivey and Scott P. Johnson for invaluable assistance with all phases of this project; Jonathan Slemmer for programming help; Myque Harris and the undergraduates in the Cornell Baby Lab for assistance with recruitment of our infant participants; Kola Ijaola and John Woschinko for help running the adult portion of this research; and Brian Scholl, whose comments greatly improved the manuscript. We are especially grateful for the contributions of the infants and parents who participated in the studies.

The stimuli used in the experiments, as well as sample eye tracking recordings from infants, can be viewed at <http://psychology.stanford.edu/~kirkham/babysquares.html>.

Correspondence concerning the infant studies reported in this article should be addressed to Natasha Z. Kirkham, who is now at the Department of Psychology, Stanford University, Building 420, Jordan Hall, Stanford, CA 94305. Correspondence concerning the adult study reported in this article should be addressed to Daniel C. Richardson, who is now at the Department of Psychology, Stanford University, Building 420, Jordan Hall, Stanford, CA 94305. E-mail: kirkham@psych.stanford.edu or richardson@psych.stanford.edu

comes of most use in navigating and learning about a multimodal changing environment.

We have studied spatial indexing by tracking the eye movements of participants, examining under what circumstances they refixate locations associated with certain objects or events. This methodology has the advantage of allowing us to investigate the role of spatial indexing behavior in adults, when they are engaged in a nonspatial task, and in very young infants who are still learning about locations and objects.

In our paradigm, participants were presented with multimodal events in certain locations, or ports, on a computer screen. An auditory stimulus relating to one of the events was presented, and we measured eye movements to the (now empty) port. Our results attest to the importance of dynamic spatial indexing of multimodal events, because we find that it is present early in development, by 6 months of age, and continues to influence adult behavior in tasks in which location information is irrelevant.

We begin by discussing in more detail the components of dynamic spatial indexing of multimodal events: encoding object location, binding multimodal properties, and indexing information. Researchers have increasingly acknowledged that theories of adult and infant cognition can and should be mutually informative (e.g., Leslie, Xu, Tremoulet, & Scholl, 1998). Therefore, we present findings from the adult and infant literature in parallel throughout this article and then proceed to test our hypotheses with both infant and adult participants.

Attending to and Encoding Objects and Locations

If attention is a process of selection, then one can ask, what is it that is being selected, and how (Kahneman & Henik, 1981)? Is attention allocated primarily to a location (Posner, 1980), a feature or set of features (Treisman, 1988), or a (preattentively defined) object (Duncan, 1984)? In other words, what is the representational basis (Egeth & Yantis, 1997), and what are the units of attention (Scholl, 2001)? Early theories of attention used spatial metaphors, such as a “spotlight” to describe how attention is deployed (Eriksen & Eriksen, 1974). Generally, orienting attention to a location aids processing of stimuli at that location (Posner, 1980). Indeed, modality-specific processing in a particular spatial location is facilitated even if a different sensory modality orientated attention to that location, regardless of whether that orientating is voluntary (Spence & Driver, 1996) or involuntary (McDonald & Ward, 2000). Although location is certainly important, there is good reason to believe that the spotlight metaphor does not fully capture how attention can be deployed. Recent functional magnetic resonance imaging (fMRI) evidence suggests that in addition to attentional processes that select a region of space, constraints are imposed on selection of objects (Arrington, Carr, Mayer, & Rao, 2000). Indeed, behavioral studies show that both adults (Duncan, 1984) and infants (Johnson & Gilmore, 1998) can allocate attention to objects in the world rather than locations in space (for a review, see Scholl, 2001). If one part of an object is cued, then attention appears to “spread” across the object, such that processing of its other features is enhanced, rather than features that are closer in space but part of other objects (Baylis & Driver, 1993; Egly, Driver, & Rafal, 1994). Johnson and Gilmore (1998) have shown that infants as young as 8 months of age demonstrate object-based attention.

Using a change detection paradigm, Luck and Vogel (1997) demonstrated that the capacity of visual short-term memory is best counted in objects rather than features. Recent work suggests that a weaker version of this hypothesis is more viable (Olson & Jiang, 2002; Wheeler & Treisman, 2002; Y. Xu, 2002); however, it remains the case that one aspect of visual working memory stores objects and is limited to small number of such representations.

Moreover, objects can be selectively attended to even when they occupy the same location in space. Early “selective looking” paradigms projected stimuli on top of each other and found that attention could be directed to one target instead of another (Neisser & Becklen, 1975). Recent behavioral (Blaser, Pylyshyn, & Holcombe, 2000; Simons & Chabris, 1999) and fMRI (O’Craven, Downing, & Kanwisher, 1999) evidence has shown further that attending to the attribute of one object enhances processing of other attributes of that object, even when different objects occupy the same location in space. Conversely, other results in the literature show that if the same object moves to different locations in space, then attentional effects such as “preview benefits” (Kahneman, Treisman, & Gibbs, 1992) and “inhibition of return” (Tipper, Driver, & Weaver, 1991), remain associated with that object. Thus in a variety of ways, it has been shown that attention can select an object over an absolute location.

Once an object is selected, however, location information continues to play a role. Chun and Jiang (1998) showed that implicit learning of spatial contexts could guide attention in a visual search task. Although it may be too strong to claim that location information is encoded “automatically” (Naveh-Benjamin, 1988) when a stimulus is attended, a range of memory studies suggest that if any property of an object has been encoded, then it is likely that its location has also been encoded (Andrade & Meudell, 1993; Ellis, 1991; Epstein & Kanwisher, 1999; Jiang, Olson, & Chun, 2000; Simons, 1996; Van der Heijden, 1993) and that location may even play a role in accessing other properties from memory (Brouwer & Van der Heijden, 1997; Johnston & Pashler, 1990; Nissen, 1985). Köhler, Moscovitch, and Melo (2001) found that naming objects improved memory for their absolute location, more so than a task judging the spatial relations between objects, suggesting that the process of object identification causes identity and location information to be bound together.

Adults, it seems, are able to flexibly deploy attention and encode spatial information with ease. But when do these abilities emerge, and at what stage might they be able to support dynamic spatial indexing? Although adults represent locations using many different spatial reference frames (J. F. Stein, 1992), infants of 4 months of age generally represent location relative to their current eye position (Gilmore & Johnson, 1997). Locations encoded in this retinocentric reference frame become invalid once a saccade is made, making spatial indexing implausible. An egocentric reference frame combines retinal and eye position information to represent the location of the targets relative to the body and is the bare minimum required to keep track of, as well as update, spatial locations. Gilmore and Johnson (1997) presented compelling evidence that by 6 months of age, infants are beginning to use egocentric reference frames to plan saccades. They used a “double-step” saccade paradigm, in which two visual targets were flashed briefly in a dark visual field. The targets were flashed one after the other with an interval short enough that the second target had already disappeared before the saccade to the first target had

finished. In this paradigm, therefore, a retinocentric reference frame would result in an incorrect saccade to the location of the second target. Gilmore and Johnson (1997) found that a movement from predominantly retinocentric reference frames to predominantly egocentric reference frames takes place between 4 and 6 months of age. Spatial indexing, as we have described it, demands that the locations of events are encoded and refixated. At around 6 months of age, infants appear to be acquiring these basic skills.

Multimodal Events

During the first 6 months of life, infants develop many intersensory capacities, which allow them to perceive correlations across modalities (for a review of this literature, see Lewkowicz, 2000). For example, newborns bind a visual stimulus with an auditory stimulus to the extent that they expect the sound to move with the associated object (Morrongiello, Fenwick, & Chance, 1998). This capacity develops to such a degree that by 6 months of age infants require relatively precise spatial collocation to equate visual with auditory stimuli (Fenwick & Morrongiello, 1998). Lewkowicz and Turkewitz (1980) showed that very young infants (3 weeks of age) use stimulus intensity as a way of binding sound and vision. By 4 months of age, infants not only perceive the bimodal nature of objects (Spelke, 1979, 1981) but also perceive speech bimodally (Kuhl & Meltzoff, 1982). Infants of 5 months exhibit the McGurk effect (e.g., the perception of /da/ when listening to the sound /ba/ paired with the visual input of a person saying /ga/; McGurk & MacDonald, 1976; Rosenblum, Schmuckler, & Johnson, 1997). In addition, infants of 5 months, when habituated to a bimodal presentation of rhythm (e.g., an audiovisual movie of a hammer tapping out a rhythm), dis-habituate to a unimodal presentation of a novel rhythm (e.g., just the visual of a hammer tapping, without the sound; Bahrick & Lickliter, 2000). By 5–7 months, infants can match faces with voices on the basis of the age, gender, and affective expression of the speaker (Bahrick, Netto, & Hernandez-Reif, 1998).

It is clear, therefore, that 6-month-olds can reliably perceive cross-modal co-occurrences, and use auditory information to disambiguate visual events (e.g., Scheier, Lewkowicz, & Shimojo, 2003). Certainly, it is true that a great deal of people's perceptual understanding (as well as a great deal of our responses to the environment) requires correctly correlating events across modalities. Gibson (1969) suggested that the responsiveness to invariant intersensory relations is a necessary part of the development of perception and learning. Lewkowicz (2000) suggested that this responsiveness becomes increasingly more complex throughout infancy, meaning that each perceived intersensory relation (e.g., sensitivity to offset–onset synchrony in an audiovisual event) acts as a foundation for the next relation in the hierarchy (e.g., sensitivity to duration as a fundamental property of intersensory equivalence).

The adult brain is also highly sensitive to intersensory relations (Calvert, 2001). There are many demonstrations of cross-modal identification: Stimuli in one modality altering the perception of another. Silent lip-reading activates auditory cortex (Calvert et al., 1997); auditory stimuli can produce the illusion of two flashes of light when only one was present (Shams, Kamitani, & Shimojo, 2000); and event-related-potential studies have shown that auditory stimulation modulates processing in the visual cortex (Shams,

Kamitani, Thompson, & Shimojo, 2001). In addition to such visual–auditory interactions, it has been shown that the tactile sensation of rubbing one's hands together can be influenced by sound (Jousmäki & Hari, 1998) and that perceived flavors can be altered by color (DuBose, Cardello, & Maller, 1980).

Cross-modal information plays a further, pivotal role in perception, which is of particular relevance to our interest in spatial indexing. As Driver and Spence (1998) point out, "A single modality alone cannot provide a stable representation of external space" (p. 254). Any modality in isolation can only indicate location relative to its own receptors; and these receptors, such as eyes or hands, can move freely in relation to each other and to external objects. These indeterminacies can be overcome only if information is integrated across modalities. A possible neural mechanism related to such integration has been found in the superior colliculus of several mammalian species (Meredith, 2002; B. E. Stein & Meredith, 1993). The "multimodal neurons" that exist in this region of the brain respond to different modalities in the same spatial register (B. E. Stein, Magalhaes-Castro, & Kruger, 1975). Because these neurons are closely connected with premotor maps in the superior colliculus (Harris, 1980), researchers have argued that they can be rapidly translated into orientating actions, such as a saccade (Frens, Van Opstal, & Van der Willigen, 1995). Indeed, behavioral results mirror the tight cross-modal integration found at the neural level. Driver, Spence, and colleagues (Spence & Driver, 1997; Spence, Nicholls, Gillespie, & Driver, 1998) have investigated the role of cross-modal information in exogenous spatial attention by presenting spatial cues and targets in different modalities. Judgments are faster for targets that occur on the cued side for all pairings of auditory, visual, and tactile stimuli appearing as either target or cue.

Orientating behaviors and spatial attention appear to be highly sensitive to the multimodal aspect of events. Moreover, given the robust and early developing intersensory capacities reviewed above, it is reasonable to assume that, by 6 months of age, infants are responsive to (if not capable of acting on) auditory–visual intersensory correspondence. Therefore, multimodal events appear to be an ideal tool to investigate spatial indexing in adults and 6-month-old infants.

Indexing of Information: Pointers, Files, and Fingers

We have seen how the locations of objects and the locations of multimodal events are encoded by infants and adults. As we argued at the outset, however, in a dynamic environment these two abilities would be most useful if combined with a third: the ability to index such information. We use the term *indexing* to refer to a cluster of phenomena in which spatiotemporal continuity is encoded and tracked in the service of a cognitive or perceptual act. We discuss claims that spatiotemporal information is used to link cognitive representations to objects in the world (Pylyshyn, 2001), to reduce working memory demands (Ballard et al., 1997), and to structure an infant's developing object concept (Leslie et al., 1998).

Pylyshyn and colleagues (Pylyshyn, 2001; Pylyshyn & Storm, 1988) have hypothesized that there is a preattentive mechanism of the visual system that functions to track several objects through space simultaneously. An index is attached to an object that has been segmented from its surroundings by a set of low-level prop-

erties (Scholl, Pylyshyn, & Feldman, 2001). The index itself does not store featural information but serves as a linking mechanism between the object in the world and the mental representation of its properties. Pylyshyn (2001) argued that an indexing mechanism provides a way to individuate and track objects, establish simple geometric relations such as “next to,” or assign sets of features. In other words, the hypothesized visual indexing mechanism helps to “situate” vision in the world (Pylyshyn, 2001).

The multiple-object-tracking paradigm was developed by Pylyshyn and colleagues (e.g., Pylyshyn, 1989; Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999) to investigate the capacities of such putative visual indexing mechanisms. In such studies, participants typically view a field of several identical two-dimensional objects. Some of these objects are identified as the targets, then all move around unpredictably. Participants are required to keep track of the targets among the distractors and to identify them later. There appears to be a strong effect of target set size: Participants can reliably track four or five targets, but performance rapidly deteriorates at higher set sizes (Pylyshyn & Storm, 1988). This capacity is largely unaffected if the objects disappear temporarily behind occluders (Scholl & Pylyshyn, 1999; although see Slemmer & Johnson, 2002). These results suggest that some form of mechanism is able to index and track at least four items robustly purely on the basis of spatiotemporal continuity (given that most of the cues that would be used in real world situations have been removed in this paradigm).

Studies of multiple-object tracking typically control eye movements by requiring participants to fixate a central cross for the duration of the task. Another form of visual indexing, however, can be seen by studying the eye movements of participants engaged in a more natural task of moving real objects around in the world. Ballard, Hayhoe, and Pelz (1995) gave participants the job of constructing a pattern of colored blocks from a model. One method participants could use was to look at the model area and memorize the pattern; each block in turn could be located in the resource area and then placed in the workspace. A second method, which is a less memory-intensive choice, would be to remember the color and location of one block from the model, collect it from the resource, place it in the workspace, and then consult the model again for the next block. The strategy used by participants, however, most often entailed the minimal possible memory demands. Participants would commonly fixate the model, pick up a correctly colored block from the resource area, fixate the model again, and then place the block in the workspace. Thus two fixations per block were made on the model—one to pick up color information, one to find its place in the pattern. This is a strategy of indexing, whereby just the location of an object is maintained in working memory and other properties can be “looked up” as they are needed (Ballard et al., 1997).

Further eye-tracking work demonstrated an interesting result of this interplay between indexing and memory. Spivey and Geng (2001, Experiment 2) demonstrated that participants systematically looked at a blank region of a grid when attempting to recall properties of an object that previously occupied that location. In a series of experiments, Richardson and Spivey (2000) extended this result to nonvisual properties, showing that even if participants were recalling not visual, but auditory, semantic information, they would saccade to empty regions of a grid that had been associated with that information.

Figure 1 shows a schematic of the experimental design used by Richardson and Spivey (2000, Experiment 1). On a computer monitor, participants watched a video clip of an actor relating a short piece of factual information, such as “Shakespeare’s first plays were romantic comedies. His last was the *Tempest*.” These talking heads appeared in turn in each of four ports of a 2×2 grid. Afterward, while looking at the blank grid, participants heard a statement that related to one of the facts (e.g., “Shakespeare’s last play was the *Tempest*”) and answered whether it was true or false. During this answer period, participants’ fixations in each of the four ports were coded. The “critical” port was the port that had previously contained the talking head that conveyed the relevant information. Richardson and Spivey found that there were almost twice as many fixations of the critical port than of each of the other ports. This result was replicated when the video clips were replaced by four identical spinning crosses (Richardson & Spivey, 2000, Experiment 2) and when the ports moved to the center of the screen during presentation (Experiment 5, watching condition).

A possible explanation for these results draws on the phenomena of context-dependent memory, in which memory is improved if the conditions that were present during encoding are reinstated during recall (e.g., Bradley, Cuthbert, & Lang, 1988; Godden & Baddeley, 1975; Winograd & Church, 1988). It is important to note that in all five experiments of Richardson and Spivey (2000; as well as in Spivey & Geng, 2001), participants’ accuracy in answering the factual question never increased as a function of fixations to the critical port. This suggests that the results cannot simply be explained as an effect of context-dependent memory. Of course, one cannot make strong conclusions on the basis of null results. Although Richardson and Spivey (2000) did not find that accuracy covaried with looking patterns, it is possible that different experimental conditions or different measures (such as reaction times) might find that memory performances can be affected by looks to the critical port.

There are two additional reasons, however, that count against context-dependent memory as an explanation of Richardson and Spivey’s (2000) results. First, the four ports are almost identical visually, therefore looking at the critical port does not reinstantiate a visual context that is particularly distinct from that provided by the other three ports. Second, during debriefing, none of Richardson and Spivey’s participants claimed to be using a strategy of looking to the critical ports to remind themselves of the fact. In summary, participants’ refixations did little to reinstantiate conditions at recall that were present during encoding and did nothing (that Richardson & Spivey, 2000, found) to improve memory. An account based on context-dependent memory would be hard-pressed to account for these features of the behavior, and so Richardson and Spivey (2000) concluded that their participants were spatially indexing auditory semantic information.

In sum, there are many examples of what can be called “indexing” in adults, although the deeper similarities between these cases have not yet been established. Our particular interest is in how objects are indexed and tracked by their location. We refer to this as *spatial indexing*, though it could also be called “object indexing” or “event indexing,” to emphasize this role of spatiotemporal information. It is clear that when adults are instructed to track the location of moving objects in a field of identical distractors, they can do so robustly. In addition, when participants are not instructed to pay attention to location—when it is explicitly irrelevant to the

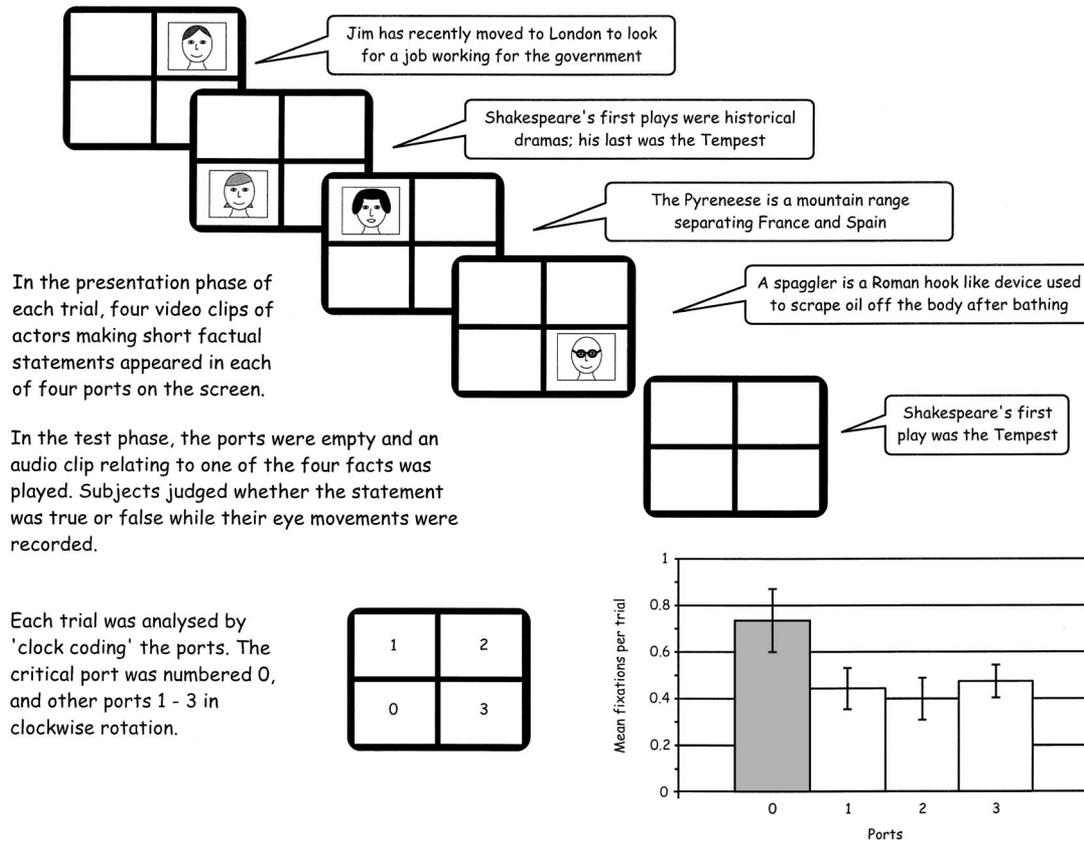


Figure 1. Design and results of Richardson and Spivey (2000), Experiment 1. Similar results were obtained when the video clips were replaced with a spinning cross (Experiment 2) and when eye movements were held in the center of the screen during presentation by a "virtual window" that the ports moved into (Experiment 5, watching condition). Error bars represent pooled standard errors. From "Representation, Space and Hollywood Squares: Looking at Things That Aren't There Anymore," by D. C. Richardson and M. J. Spivey, 2000, *Cognition*, 76, pp. 275 and 276. Copyright 2000 by Elsevier Science. Adapted with permission.

task of remembering spoken facts—adults continue to spatially index information and systematically fixate empty locations. In this way, spatial indexing occurs across different paradigms and task demands as a prevalent feature of adult behavior.

It has been argued that indexing, or tracking spatiotemporal information, plays a pivotal role in an infant's developing "object concept." Leslie et al. (1998) recently proposed a theory that drew together aspects of Pylyshyn's (1989, 2001) theory of visual indexing, and Kahneman and Treisman and colleague's (Treisman, 1988; Kahneman et al., 1992) object file theory. In their account, infants have object indexes: mental tokens that function as pointers to objects and track their location. Initially, objects are indexed only by location (i.e., features play little or no role in object individuation); later on in infancy, "dual-route" indexing develops, allowing indexing of objects using both featural as well as location information (Leslie et al., 1998). A series of studies by F. Xu and Carey (1996), in which objects were placed behind a screen after being presented either simultaneously to the infant, or one at a time, supports this idea by showing that 10-month-olds could individuate the objects only when they saw them at the same time, whereas 12-month-olds could individuate the objects in both conditions. Using similar methods, Tremoulet, Lee, and Leslie (1998;

cf. Kaldy & Leslie, 2003) found that 6- and 9-month-olds did not look longer when the shape of objects unexpectedly changed, suggesting that they do not bind this information to the object representation.

These studies provide elegant preliminary evidence for the primacy of spatiotemporal information as a cue to object individuation. They do not, however, address infants' ability to track these objects continually through space (e.g., as the multiple-object-tracking paradigm does for adults' tracking abilities) or to encode object locations for memory.

Dynamic Spatial Indexing of Multimodal Events by Adults and 6-Month-Olds

In the context of a dynamic environment and multiple sensory modalities, certain aspects of the perceptual system—object location encoding, cross-modal integration, and indexing—would have a particular utility when operating in concert. This cluster of abilities has not been directly investigated before. Various forms of indexing have been studied in infants and adults, but they have only used static information (Richardson & Spivey, 2000) or only presented visual stimuli (e.g., Ballard et al., 1997; Leslie et al.,

1998; Scholl & Pylyshyn, 1998). Therefore, to study dynamic spatial indexing of multimodal events, we modified Richardson and Spivey's (2000) paradigm such that participants were presented with multimodal events and required to encode and track moving locations.

We hypothesized that a propensity for spatial indexing is not just a feature of the mature adult visual system but emerges by 6 months along with some of the first uses of adult-like spatial reference frames (Gilmore & Johnson, 1997). To investigate spatial indexing in infants, we developed a simplified version of Richardson and Spivey's (2000) paradigm. Infants were shown movies of brightly colored toys that moved in time to a sound. Test phases consisted of the auditory element of the movie alone, and infants' looking times to empty ports on the screen were measured.

With these parallel adult and infant paradigms, we investigated three empirical questions. First, how dynamically can adults index factual, auditory information? Specifically, if adults associate information with a port on a computer screen, will they still make a saccade to that empty port if it moves around the screen? Second, if infants are sensitive to a wide range of perceptual cues, will they saccade to a blank region of a computer screen when they hear a sound that was previously associated with a multimodal event occurring in that location? Third, will infants be able to bind a multimodal event to a port that moves around the screen prior to the test phase?

Experiment 1

Richardson and Spivey (2000) presented evidence that when their participants indexed semantic facts, they did not use retinotopic oculomotor coordinates or absolute locations in space. It is not clear, however, what frame(s) of reference spatial indexing does use or whether locations can be spatially updated independently of one another. In the adult literature there are several examples of attentional effects adhering to one or more objects as they move through space (Scholl & Pylyshyn, 1999; Tipper et al., 1991). For example, in Kahneman et al.'s (1992) "reviewing paradigm," participants first saw a preview field, consisting of a number of boxes, each of which could contain a letter. Later, participants saw a target field consisting of the same number of boxes and a single letter, which was to be named as quickly as possible. Kahneman et al. found an object-specific preview advantage: Identification was facilitated if the target letter had been seen in the same box in the preview field. This effect was found even if the boxes changed location by smoothly animating between the preview and target fields. These results suggest that the effect of letter identity was not bound to an absolute location but moved with its associated box. Therefore, simple visual properties can be bound to a moving object.

The current experiment incorporates elements of Kahneman et al.'s (1992) study with the Richardson and Spivey (2000) paradigm to investigate whether complex, auditory, semantic information can be bound to a moving object. Rather than using a 2×2 grid of ports that remained in one location or moved around as a whole, in the current experiment, two separate ports, or boxes, were used. A spinning cross appeared in each of these ports while the participant heard a short fact. Then, before the question period, the ports either stayed in place, moved clockwise or moved counterclockwise (see Figure 2). In the following analyses, the "criti-

cal" port (or critical location) refers to the port that was previously associated with the fact that is questioned. During the question period, fixations were counted in the critical and noncritical ports. Following Richardson and Spivey (2000) and Kahneman et al. (1992), we hypothesized that regardless of any movements the ports make, during the question period there will be more fixations in the critical port.

Method

Participants. Twenty-two undergraduate students participated in exchange for course credit. All had vision that was normal or corrected-to-normal by soft contact lenses.

Apparatus. Eye movements were monitored by an ISCAN (Burlington, MA) eyetracker mounted on top of a lightweight headband. The eye camera provided an infrared image of the left eye sampled at 60 Hz. The center of the pupil and the corneal reflection were tracked to determine the direction of the eye relative to the head. A scene camera, yoked with the view of the tracked eye, provided an image of the participant's field of view. Gaze position (indicated by crosshairs) was superimposed over the scene camera image and recorded onto a Hi8 (Sony) VCR with 30-Hz frame-by-frame playback. Accuracy of the gaze position record was approximately 0.5° visual angle. A microphone picked up the auditory stimuli and the participants' verbal responses, and the video record was synchronized with the audio record for all data analysis. The calibration grid and stimuli were presented on a Macintosh 7200/90 computer with a 20-in. (50.8-cm) monitor and external speakers.

Stimuli. Sixty-four short statements were constructed. The statements either were factual (e.g., "In Fiji, 46% of the population is under 14, and the life expectancy is about 40 years") or concerned fictional characters (e.g., "Try as he might, Gary could not stop himself from alphabetizing other people's CD collections"). For each fact, a question was constructed. For half of these the correct answer was "no" (e.g., "Is 14% of the population of Fiji over 46?"), and for the other half the answer was "yes" (e.g., "Did Gary have an obsessive habit?"). The statements and questions were recorded by an experimenter and converted into mp3 sound files. All auditory stimuli were recorded in mono and then played at equal volume through two speakers on either side of the computer monitor.

A spinning red cross was presented while participants listened to the facts. The cross subtended approximately 6° of visual angle. The cross appeared in one of two ports placed in the left and right halves of the screen. Each port was a black box frame subtending approximately 9° of visual angle.

Design. To mask our hypothesis, we asked participants to sign a consent form that suggested that the study was concerned with "how fixing the eyes on a certain point might affect memory for facts." They were instructed to fixate the spinning cross whenever it appeared, and it was implied that the eye tracker was used solely to ensure compliance. The eye tracker was then calibrated by using a 3×3 grid of red dots on the computer monitor. After calibration, participants completed 32 trials of the experiment.

Figure 2 shows a schematic of a trial. First, the two ports appeared on the left and right halves of the screen. At random, a fact sound file was then played, and for the duration, the spinning cross appeared in one of the ports. This was repeated by using another fact and the other port. Then participants saw one of three movement conditions. For 3 s, the ports either remained in place or moved into vertical alignment in the top and bottom halves of the screens. The ports took a circular path and moved in either a clockwise or a counterclockwise direction while maintaining their local orientation. In 12 of the 32 trials, the ports remained in position; in 10, they moved clockwise; and in another 10, they moved counterclockwise.

With the two ports visible, and no spinning cross present, a question sound file was played. The question related to one of the previous two facts, selected at random. Participants held the mouse in their hand, and

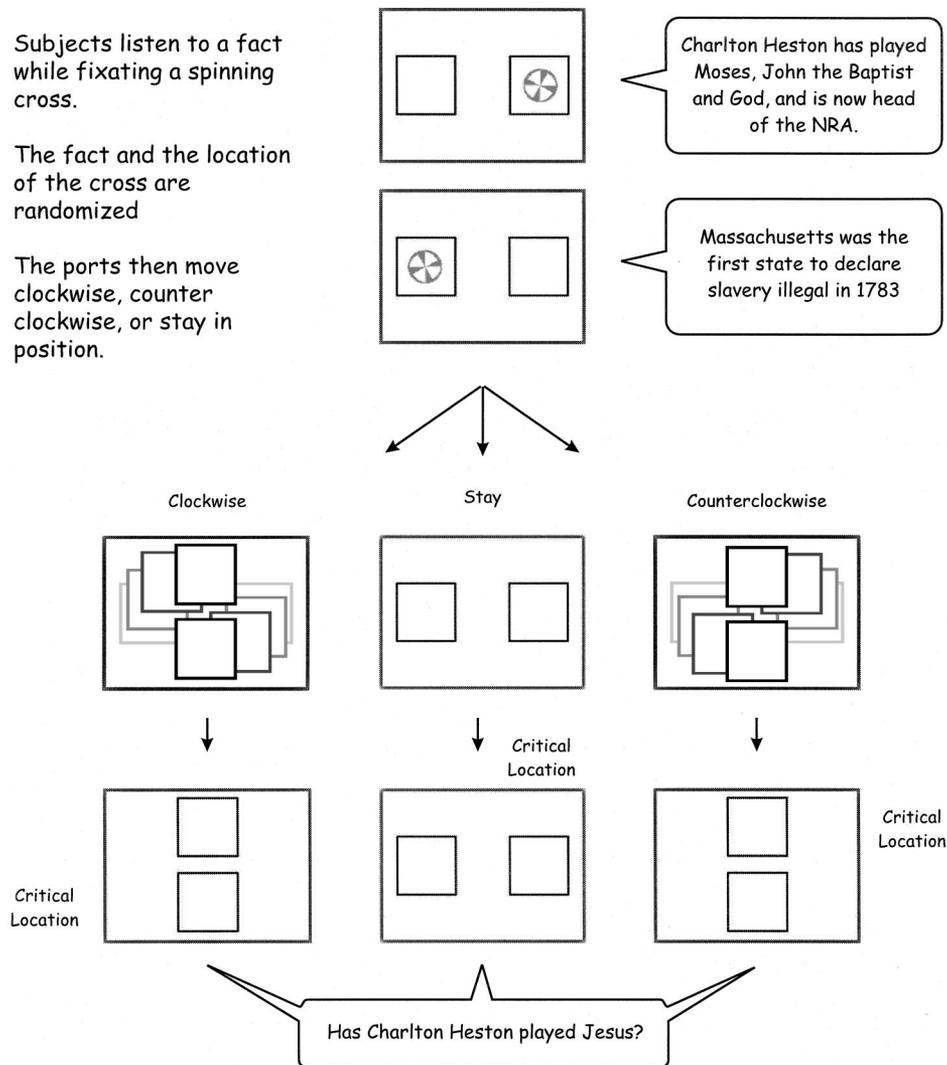


Figure 2. Schematic of Experiment 1. NRA = National Rifle Association.

after answering “yes” or “no” out loud, clicked the mouse button to start the next trial.

Results

Participants’ eye movements during the fact presentation periods were checked to ensure that participants complied with instructions and fixated the red cross whenever it appeared. Debriefing sessions indicated that participants either had no hypothesis or guessed that the eye tracker was only used for this purpose. Therefore, they were not aware that their eye movements during the question period were of interest to the experimenters.

The participants’ eye movements were analyzed in the period beginning from the onset of the question and ending with the participants’ response. The coder did not hear which question had been associated with which fact, and so did not know which port was the critical location. The number of fixations in each port was counted. A fixation was operationalized as steady eye-position crosshairs for 100 ms or more (three frames on the video tape). The

criterion of 100 ms was chosen to exclude crosshair movement that was due to blinking or saccades across a port. The participants’ verbal responses were also coded and are discussed separately in the *Accuracy analyses* subsection below.

Fixation analyses. A 2 (port type: critical vs. noncritical) \times 3 (movement condition: stay vs. clockwise vs. counterclockwise) analysis of variance (ANOVA) on fixations revealed a highly significant effect of port type, $F(1, 21) = 27.56$, $MSE = 0.425$, $p < .001$. The critical ports received more than twice the number of looks than the noncritical port (mean fixations: critical = 1.10, noncritical = 0.50). Effect size analysis (Cohen, 1988) shows that this port type effect was large ($d = 1.02$). There was no main effect of movement type ($F < 1$). In addition, there was no interaction between port and movement conditions ($F < 1$), suggesting that regardless of how they moved, the critical ports were looked at far more often (see Figure 3).

Using a separate 2 (port type: critical vs. noncritical) \times 4 (absolute location: left vs. right vs. top vs. bottom) ANOVA, we

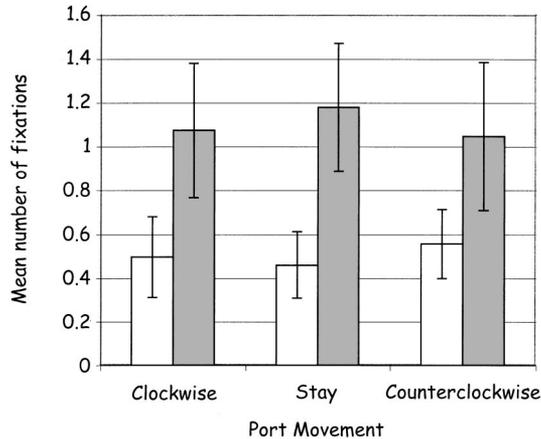


Figure 3. Experiment 1 results. Error bars represent 95% confidence intervals. The effect of port type was significant ($p < .001$). Gray bars indicate critical port type; white bars indicate noncritical port type.

found no main effect of the absolute location of the ports ($F < 1$). There was, however, a significant interaction between port type and absolute port location, $F(3, 63) = 3.04$, $MSE = 0.192$, $p < .05$. The difference between critical and noncritical ports was greater for the left (0.85) and right (0.73) positions than for the top (0.35) and bottom (0.48) positions. This interaction with absolute screen location does not impact our hypothesis, however, because simple effects analyses show that in every position, the critical ports were always looked at significantly more often: left, $F(1, 21) = 22.82$, $MSE = 0.351$, $p < .001$; right, $F(1, 21) = 16.92$, $MSE = 0.345$, $p < .001$; top, $F(1, 21) = 10.18$, $MSE = 0.131$, $p < .01$; bottom, $F(1, 21) = 15.08$, $MSE = 0.167$, $p < .01$.

A final analysis examined the effect of recency. The data were coded according to whether the question referred to the fact that was heard first (nonrecent) or last (recent) during the presentation phase. A 2 (port type: critical vs. noncritical) \times 2 (recency: recent vs. nonrecent) ANOVA revealed no main effect of recency ($F < 1$). Moreover, there was no interaction between recency and port type ($F < 1$), showing that regardless of whether participants were questioned about a fact that they had just heard or one they had heard previously, they looked more often at the critical port.

Accuracy analyses. Participants gave the correct verbal response on an average of 77% of trials. The accuracy of participants' answers was compared with their looking behavior by comparing trials in which no looks were made to the critical port with trials in which one or more looks were made to the critical port. Three participants had to be removed from this analysis because they did not contribute to every cell of the design. A 3 (movement condition: stay vs. clockwise vs. counterclockwise) \times 2 (critical look: no looks vs. 1 or more looks) ANOVA was performed on the accuracy data of the remaining participants, and revealed no significant main effects. Accuracy did not seem to be affected by the movement of the ports (clockwise = 79%, counterclockwise = 74%, stay = 77%; $F < 1$). Moreover, if the critical port was looked at (76%) or not looked at (77%), accuracy did not differ substantially ($F < 1$). There was, however, a significant interaction between movement condition and critical looks, $F(2, 36) = 3.33$, $MSE = 0.045$, $p < .05$. This interaction was driven by

one anomalous cell (trials in which there was a counterclockwise movement, and a look to the critical port), which had an unusually low accuracy level of 66% and significantly differed from some of the other cells (Fisher's least significant difference = 14%). The remaining five cells had accuracy levels between 75% and 83%, which did not differ significantly from each other.

Accuracy was also analyzed with a 4 (absolute location: left vs. right vs. top vs. bottom) \times 2 (critical look: no looks vs. 1 or more looks) ANOVA. Six participants had to be removed from this analysis for not contributing to all cells in the design. The main effect of absolute location on accuracy was not significant ($F < 1$), nor was the main effect of critical look ($F < 1$). There was no reliable interaction between absolute location and critical look, $F(3, 45) = 1.49$, $p = .23$.

Discussion

Richardson and Spivey's (2000) main finding was replicated in this experiment: When answering a question, participants had a strong tendency to fixate the empty port that had previously been associated with the relevant factual information. This experiment further demonstrated that the effect held even if the ports occupied different locations during the presentation and test periods. Indeed, regardless of whether the ports moved or stayed still, participants were about twice as likely to look to the critical port than the noncritical port. As in several previous experiments (Richardson & Spivey, 2000; Spivey & Geng, 2001), participants were not more or less likely to get the factual question correct if they looked at the critical port. Moreover, in this experiment the two ports were visually identical squares that were in different locations during fact encoding and retrieval, providing stronger evidence against a contextual memory explanation. We conclude that spatial indexing can associate complex, semantic information to a moving object, in a manner analogous to the "previewing effect" of letter identity binding to objects (Kahneman et al., 1992).

Experiment 2

To examine infants' spatial indexing abilities, we tested 6-month-old infants by using a simplified version of Experiment 1, in which the ports remained stationary. Infants were presented with two bimodal cues (visual and auditory stimuli) and then their memory of an amodal attribute (spatial location) was probed by playing the sound previously associated with the visual stimulus. As in Experiment 1, during the test trials, the critical location was defined as the port which had previously been associated with the auditory stimuli. Similarly, our hypothesis was that infants would look more at the critical location, even though both of the ports on screen were empty during the test trials.

Method

Participants. Eleven full-term 6-month-old infants (7 female, 4 male) composed the final sample (age: $M = 6.24$ months, $SD = 0.43$). Five additional infants were observed but not included in the analyses because of fussiness ($n = 2$) or poor calibration of point of gaze (POG; $n = 3$). The infants were recruited by letter and telephone from hospital records and birth announcements in the local newspaper. Parents and infants received a small gift (a baby T-shirt or toy) for their participation.

Apparatus. A Macintosh G4 computer and 76-cm color monitor were used to present stimuli. The infants were shown the stimulus displays as their eye movements were recorded. An Applied Science Laboratories (Bedford, MA) Model 504 corneal reflection eye tracking system was used to collect eye movement data. A remote pupil camera with a pan/tilt base was placed on the table below the stimulus monitor. The stimulus viewed by the infant was imported directly into the eye tracker computer from the Macintosh. The eye tracker also fed a signal into a videotape recorder in the form of crosshairs superimposed on the stimulus for purposes of off-line data coding (see the *Results* section). Figure 4 shows the layout of the laboratory.

Procedure. Infants were tested individually while seated in a caregiver's lap 100 cm from the stimulus monitor. Each session required two experimenters, an "observer" and a "video experimenter." The observer watched the infant through a peephole in one of the two partitions extending out from either side of the stimulus monitor and held a remote control that directed the pupil camera. The video experimenter sat behind the stimulus monitor and watched an image of the infant's pupil on a 25-cm achromatic monitor, and the POG and stimulus on the VTR monitor. Both the observer and the video experimenter were out of sight of the infant (see Figure 4). Before the stimulus presentation, the room lights were turned off and the infant was shown an Ernie and Bert cartoon to engage his or her attention as the observer directed the pupil camera toward the infant's eye with the remote control. After the left eye of the infant was in view, the video experimenter changed from this "manual" mode of camera control to an "automatic" mode, during which the camera remained directed at the pupil despite small displacements of the infant's head (via an algorithm built into the eye tracker). (Occasionally during the experiment, the infant moved his or her head more quickly than the camera could follow, such that the pupil was lost from view. At this point the video experimenter changed from the automatic mode back to manual, the observer again located the pupil in the camera, and automatic control was resumed.)

Following acquisition of the pupil image, and as the infant watched the cartoon, adjustments were made on the eye tracker to maximize robustness of the POG. This varied somewhat from infant to infant with respect to reflectance of infrared and visible illumination (corneal and pupil reflection, respectively). The infant was then shown a dynamic flashing circle to attract his or her attention (the "attention-getter"); this attention-getter was used to calibrate the infant. The eye tracker was calibrated on each infant's

left eye by using a two-point calibration routine (i.e., the POG for upper left and lower right locations were used, and other locations were interpolated by the computer) and then checked against two different points.

If the calibration was satisfactory, the first block of the experiment was begun. Each block consisted of six presentation trials and two test trials (see below). Blocks were presented repeatedly until the infant lost interest in the stimuli and did not return his or her gaze to the screen when the attention-getter appeared.

Stimuli. The ports consisted of two boxes with white outlines that were presented on a black background. The boxes subtended approximately 11.4° of visual angle and each were centered approximately 10.3° from the midline. When required, the attention-getter appeared in the center of the screen, subtending 5.2° at its fullest extent.

The stimuli consisted of two Quicktime movies of objects moving in time to repeated sounds. For each infant, two objects were randomly chosen from a set of four (a rattle, a toy cat, a toy dog, and a toy duck). Each object subtended approximately 8° of visual angle. The objects were randomly allocated a sound each from a set of four. The sounds differed from each other in their rhythm, tone, and pitch changes. These elements were combined in two 8-s Quicktime movies, during which the objects moved within their port in time with their sounds. For example, one sound was a telephone-like "Brring!" that occurred once a second—in time to this, the object vibrated. Another sound was a melodic "de dump" sound that occurred twice a second, during which the objects rotated 45° back and forth to one side and then the other. When creating the stimuli, we ensured that (a) in each event visual and auditory elements were synchronous and (b) the two events were highly distinct from one another. The events were randomly assigned to the left and right ports and appeared only in those locations for the entire duration of the experiment.

Design. Figure 5 presents a schematic of the experimental design. Each block consisted of six presentation trials and two test trials. Each trial began with a centrally placed attention-getter that flashed on and off in time to siren-like sound. This remained on screen until the video experimenter pressed a button to signal that the infant had fixated it.

The six presentation trials were made up of three pairs of trials. Each pair consisted of presentations of each of the two multimodal events in their respective locations. Order of presentation within the pairs was randomized. This constraint meant that across the six presentation trials, the order of presentation was pseudorandomized: The infant never saw the same event more than twice in succession.

The six presentation trials were followed by two test trials, interspersed once more with the attention-getter, so that the test trials began with the infant fixating the center of the screen. In each test trial, the infant was presented with the auditory element of one of the multimodal events. This sound lasted for 8 s, during which the infant viewed two empty ports. After gaze was returned to center again by the attention-getter, the empty ports display was brought back and the auditory element of the other event was presented for 8 s.

The block of six presentation trials and two test trials was repeated. The experiment was an infant-controlled procedure and stopped when the infant no longer responded to the attention-getter.

Results

Figure 6 shows an example scanpath of an infant in two test trials of our experiment. In this case, one can see that the majority of the infant's fixations are to the critical side of the screen. This trial is a particularly striking example of the looking behavior we hypothesized. To demonstrate that other infants showed a similar effect, we measured the total amount of time the infants spent looking to each side of the screen. Looking times (rather than the fixation counts used in Experiment 1) were used as the dependent variable, because infant eye position signals can be noisier than those from adults. If the track was jittery, a fixation was not always

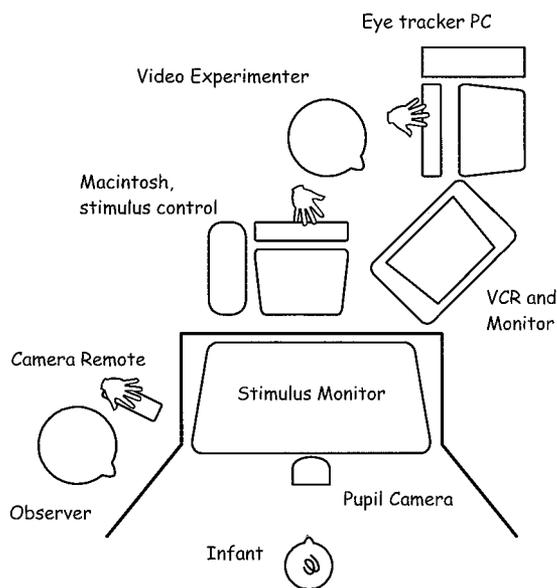


Figure 4. Laboratory setup for Experiments 2 and 3.

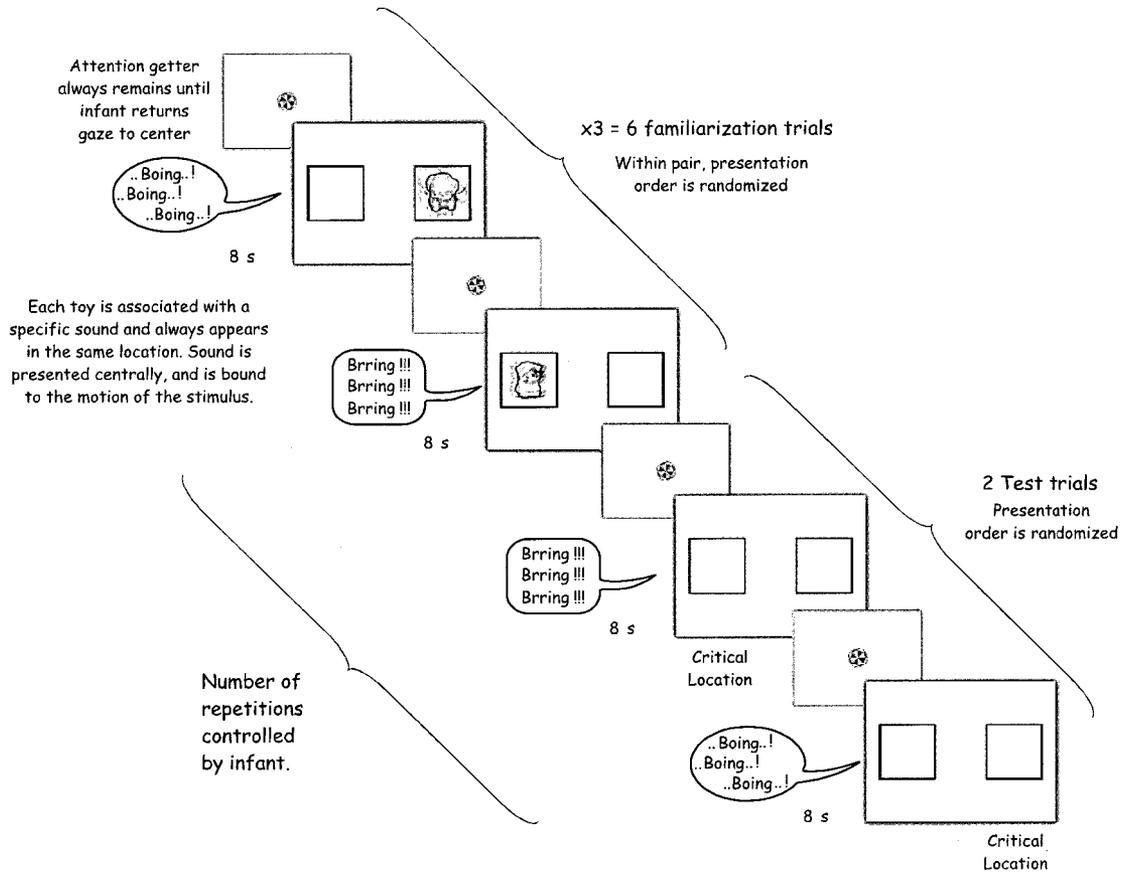


Figure 5. Schematic of Experiment 2. On the computer screen, the background was black, box lines were white, and stimuli were in full color.

registered by the software, but the side of the screen that was fixated could easily be discriminated from the recording.

Infants completed between 2 and 10 test trials ($Mdn = 6$). Each test trial was coded blind from video tape by two coders. A divider was placed down the center of the display. Using stopwatches, one coder measured the total length of time during the test trial that the crosshairs appeared on the left half of the screen, and another coder did the same for the right half of the screen. The volume on the display was muted, so that the coders could not hear what sound was being played and hence did not know which was the critical location.

A 2 (location: left vs. right) \times 2 (port type: critical vs. noncritical) repeated measures ANOVA revealed a significant main effect of port type, $F(1, 10) = 5.84$, $MSE = 0.746$, $p < .05$. During test trials, infants looked longer at the critical location than at the noncritical location ($M = 2.77$ s vs. $M = 2.14$ s, respectively; see Figure 7). Effect size analysis shows that this is a large effect ($d = 0.88$). There was no effect of absolute location; looking time to either side of the screen was not reliably different (M , left = 2.65 s; M , right = 2.26 s; $F < 1$). Although the effect of port type was numerically larger when the test trial was on the left rather than the right, this interaction between location and port type was not significant, $F(1, 10) = 4.22$, $p = .07$.

Possible recency effects were tested by coding a trial as "recent" if the infant was tested on a sound that had been part of the

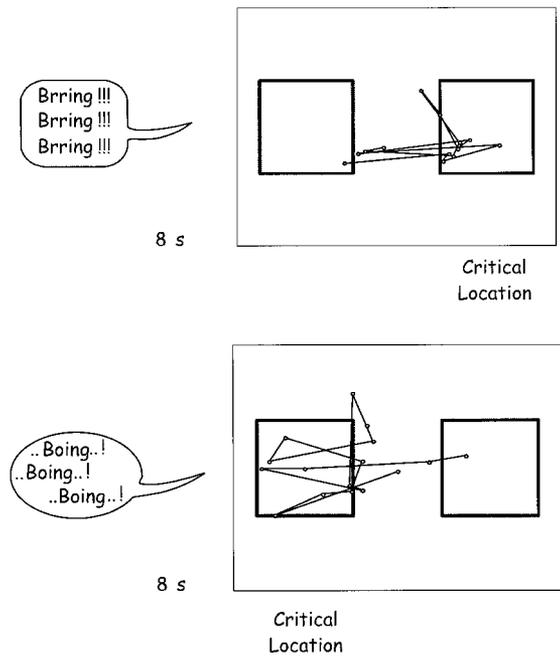


Figure 6. Scanpath of a 6-month-old infant during two test trials of Experiment 2. Circles represent fixations.

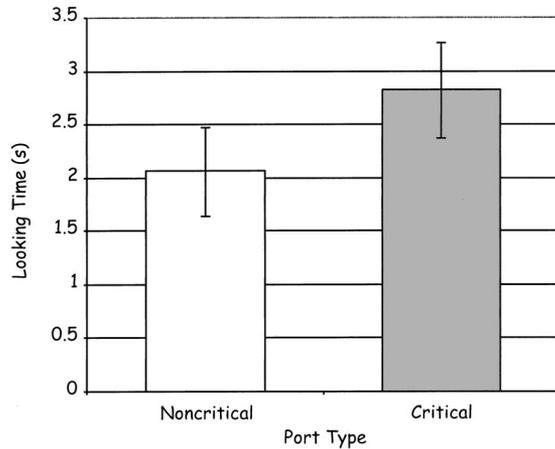


Figure 7. Experiment 2 results. Error bars represent 95% confidence intervals. The effect of port type was significant ($p < .05$).

immediately prior presentation trial and as “nonrecent” if there had been other intervening trials. A 2 (recency: recent vs. nonrecent) \times 2 (port type: critical vs. noncritical) repeated measures ANOVA showed no significant main effect of recency, $F(1, 10) = 2.50$, $p = .16$, and no interaction between recency and port type ($F < 1$). Therefore, infants looked longer at the critical port during the test trials regardless of whether the sound had been part of the previous presentation trial or whether there had been other sounds presented in between.

Discussion

Infants of 6 months of age looked reliably longer to the critical side of the screen than to the noncritical side of the screen. These results suggest that the infants had (a) bound the toy and its associated sound together such that they anticipated co-occurrence of the two and (b) bound both the events to a specific location such that on hearing the associated sound, they looked significantly longer at that specific location. One interpretation of this behavior is that the infants were predicting a visual event to occur in that location.

This result concurs with Von Hofsten’s (1980) finding that, at the same age, babies exhibit anticipatory reaching to moving objects and with Clifton, Muir, Ashmead, and Clarkson’s (1993) finding that, at around 6 months of age, infants start to use an auditory cue to reach toward an object in the dark. Experiment 2 demonstrates, therefore, that infants can associate a multimodal event with a location and access it with an auditory cue. This is just one step short of the spatial indexing shown by our adult participants, which occurred even if the ports changed location.

Experiment 3

This experiment was designed to test the flexibility of spatial indexing in 6-month-olds by moving the blank ports from vertical to horizontal alignments after familiarization. We predicted that, like adults, infants’ spatial indexing would not be disrupted by movement of the ports, and that they would make more saccades to the critical location during test trials.

Method

The apparatus and procedure were identical to Experiment 2.

Participants. Twelve full-term 6-month-old infants (6 female, 6 male) composed the final sample (age: $M = 5.92$ months, $SD = 0.28$). Six additional infants were observed but not included in the analyses because of fussiness ($n = 3$) or poor calibration of POG ($n = 3$). The infants were recruited by letter and telephone from hospital records and birth announcements in the local newspaper. Parents and infants received a small gift (a baby T-shirt or toy) for their participation.

Stimuli. The stimuli were the same as those used in Experiment 2, but to accommodate the movement of the ports, their on-screen size was reduced. The attention-getter subtended 4° of visual angle, and the ports subtended 8.6° and were spaced 8° from the center of the screen in both their vertical and horizontal alignments.

Design. The design of this experiment was identical to Experiment 2, apart from the changes to the presentation trials described here. At the start of the block, the two ports were aligned vertically on the screen, and the multimodal events occurred in the top and bottom positions. After the six presentation trials, gaze was returned to center by the attention-getter, and then the ports smoothly translated in either a clockwise or anticlockwise direction while maintaining their local orientation. During this movement, the ports maintained their orientation. This animation took 4 s. The two test trials then occurred exactly the same as in Experiment 2. Finally, the ports smoothly translated back to their original places in a vertical alignment, so that a new set of presentation trials could begin again.

Results

Infants completed between 3 and 16 test trials ($Mdn = 6$), and the data were coded in the same way as in Experiment 2, with an additional between-participants variable of direction of translation. Thus, we analyzed the data with a 2 (location: left vs. right) \times 2 (port type: critical vs. noncritical) \times 2 (translation: clockwise vs. counterclockwise) repeated measures ANOVA. As hypothesized, there was a significant main effect of port type, $F(1, 10) = 36.38$, $MSE = 0.360$, $p < .001$. Infants looked longer at the critical location ($M = 2.81$ s) than at the noncritical location ($M = 1.77$ s), as shown in Figure 8. The size of the port type effect was large ($d = 1.24$). The main effect of absolute location was not significant (M , left = 2.01; M , right = 2.57), $F(1, 10) = 3.82$, $p = .08$.

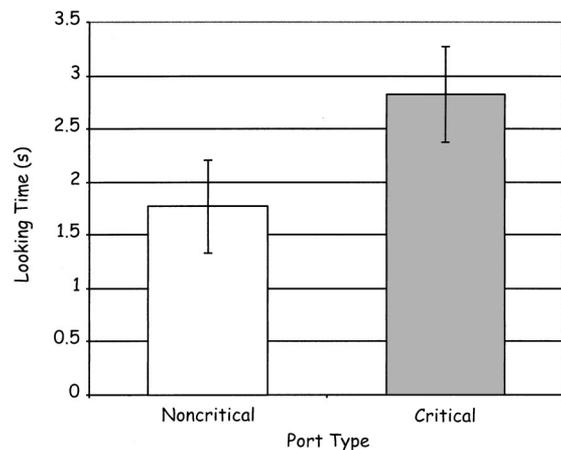


Figure 8. Experiment 3 results. Error bars represent 95% confidence intervals. The effect of port type was significant ($p < .001$).

The interaction between location and port type did not approach significance ($F < 1$).¹

There was a significant main effect of translation direction, $F(1, 10) = 9.64$, $MSE = 1.129$, $p < .05$, such that infants looked more overall when the preceding rotation had been in a clockwise (2.77 s) rather than a counterclockwise direction (1.81 s). The effect size of translation direction was large ($d = 1.26$), but it did not significantly interact with either absolute location ($F < 1$) or port type ($F < 1$), nor was there a significant three-way interaction between all the variables ($F < 1$). We investigated whether this rotation preference was in some way related to a preference the infants may have had for presentation events that occurred in the top versus the bottom of the screen. A new variable, presentation location, indicated whether at test the infant was looking at a port that (during presentation) had been on the top or on the bottom. The main effect of presentation location was not significant ($F < 1$), nor did it interact with port type (critical vs. noncritical; $F < 1$). The interaction between presentation location and rotation direction was marginally significant, $F(1, 10) = 3.822$, $MSE = 0.976$, $p = .079$. However, this possible interaction reflects the slight preference for the right side of the screen discussed above. There was no three-way interaction between presentation location, rotation direction, and port type ($F < 1$).

As in previous experiments, possible recency effects were examined by a 2 (recency: recent vs. nonrecent) \times 2 (port type: critical vs. noncritical) repeated measures ANOVA. There was no significant main effect of recency ($F < 1$) and no interaction with port type ($F < 1$).

The results of Experiment 2 and Experiment 3 were compared directly in a 2 (port type: critical vs. noncritical) \times 2 (experiment: static ports vs. moving ports) repeated measures ANOVA. As would be expected, the main effect of port type was highly significant in the comparison, $F(1, 21) = 30.78$, $MSE = 0.248$, $p < .001$. There was no main effect of experiment ($F < 1$), nor did the interaction approach significance, $F(1, 21) = 1.26$, $p = .28$.

Discussion

When the ports moved, 6-month-olds looked significantly longer at the new critical locations. In addition, not only did we find significantly longer looking time to the critical location, but the comparison between Experiments 2 and 3 showed that movement of the ports between presentation and test had no significant effect. This finding echoes the results of Experiment 1, in which adults' looks to the critical location were unaffected if the ports moved or stayed still.

The main effect of rotation is puzzling. There is evidence that from birth, infants can discriminate rotation directions (Laplante, Orr, Neville, Vorkapich, & Sasso, 1996), but there do not appear to be any reports in the literature of an overall preference for one direction over another. In the current experiment, infants who saw a clockwise rather than a counterclockwise translation prior to the test trials spent more time overall looking at the stimulus display. We suspect that because rotation direction was a between-subjects variable, and there were six infants in each condition, this difference was produced by a few outliers. Indeed, two infants in the counterclockwise condition looked at the display for less than half average time. It is important to note that the direction of the port translation did not interact with any other variables but simply

appeared to modulate the looking times across all conditions. Therefore, it does not impact our conclusion: 6-month-olds are able to associate multimodal events with ports, track these ports as they move to a different location, and fixate one of them when presented with only the auditory element of the associated event.

General Discussion

The three experiments reported here were designed to investigate the dynamic spatial indexing of multimodal events in adults and infants. Experiment 1 showed that adults continue to make saccades back to the location associated with a certain fact, regardless of whether that location has stayed still or has moved. Experiments 2 and 3 showed that infants not only appear to have the same tendency to bind a sound with a location but demonstrate adult-like flexibility in their spatial indexing. There are three features of our findings that are of particular interest, especially because these observations hold equally for the adults and 6-month-olds. First, during the test phase of all three experiments, participants were viewing two empty frames—nevertheless, we found systematic eye movements toward particular blank regions of space. Second, the eye movements were driven by a combination of the current auditory stimulus (a factual question or a rhythmic sound) and a memory of the multimodal events that had previously occurred in each port. Third, the memory of the multimodal events was bound to a moving location, not an absolute location, because the ports occupied different locations between presentation and test phases.

Our data provide evidence that three particular perceptual abilities—encoding of object location, sensitivity to cross-modal information and spatial indexing—can cooperate to drive the eye movements of adults and infants systematically around an empty display. Below we offer an interpretation of our results, drawing on the notions of the external world as a source of rich, intersensory regularities and as a potential repository for information. We then compare our account of dynamic spatial indexing of multimodal events with other related theories of indexing in adults and infants.

Multimodal Regularities and “External Memory”

In Experiment 1, adult participants appeared to spatially index spoken facts. The only information needed for the task was auditory, and participants' memory for location was never tested explicitly. Nevertheless, there was strong evidence that spatial loca-

¹ It is conceivable that during the test phase, infants might have made saccades to the pretranslation locations (i.e., the top or bottom locations where the squares had previously been seen during the familiarization phase). If infants selectively looked at the screen location where the event had previously occurred, this would demonstrate a memory for the original, absolute location of the critical event. In fact, during the test phase, infants' saccades were directed to the empty ports on the left and right of the screen, with almost no looks to the top and bottom locations. This is not surprising, given that the original locations were empty at the time and therefore less attractive than the new locations. Thus, the design of the current experiment provides scant data to answer the question of whether infants retain any representation of the original event location, in addition to tracking the new location. This interesting question will be the subject of future work.

tion was not only associated with the semantic information but also associated in such a way that an action (an eye movement) was launched when the information was accessed. However, from the limited standpoint of memory performance alone, these eye movements were epiphenomenal: Fixations in the critical port did not improve memory for the fact. How then can we explain these systematic looks to blank and unhelpful regions of space?

One interpretation draws a parallel between remembering pieces of factual information and constructing a pattern of colored blocks (Ballard et al., 1995). Eye movement patterns in the latter task indicated that participants used an indexing strategy. Rather than memorizing the properties of a given block in the model, participants appeared to fixate it every time they required information. This behavior can be described by Kirsh and Maglio's (1994) concept of "epistemic actions": physical actions that aid mental computations. They documented the surprising behavior of expert Tetris players, who make more key presses than novices when attempting to tessellate a shape. The experts have learned that the action of pressing a key to rotate a shape is faster than mentally rotating it. In other words, they gain information through an action because it is more efficient than a mental operation.

Participants' reliance on refixation, instead of memorization, in the block-moving task was demonstrated by Hayhoe, Bensinger, and Ballard (1998). They replicated the task on a computer screen and occasionally changed the color of a block during a saccade toward it. The color change was very rarely noticed by participants, suggesting that they had not previously encoded color information (see also Hayhoe, 2000; Rensink, 2002; Simons & Levin, 1997). Participants were acting as if the world were an "external memory" (Brooks, 1991; Clark, 1997; O'Regan, 1992; Spivey, Richardson, & Fitneva, *in press*), only accessing properties of the block, moment by moment, as the task required. Every system of information storage needs a system of information retrieval: Spatial indexing could be one way that the external memory of the world is accessed.

Similarly, it could still be argued that participants in Experiment 1 were acting as if the factual information they had heard could be retrieved from that external memory. When a fact was heard, participants associated the information with a port on the computer screen. When the information was needed again, during the question period, the association was activated and a saccade was launched to retrieve that information. Of course, in this case, there was no useful information there at all, and so accuracy in answering the question did not increase with fixations to the critical port.

This interpretation suggests that in terms of their looking behavior, participants were treating pieces of evanescent, auditory, semantic information as if they were physical objects in the world, there to be reinspected whenever the need arose. Why is spatial indexing so ingrained in our perceptual and cognitive processing? Perhaps it is because the primate visual system has evolved to deal with objects and locations, rather than words and facts. Our experiments with infants suggest another part of the answer: Spatial indexing emerges by 6 months of age along with some of the earliest demonstrations of mature, egocentric, spatial processing. The question then becomes, why are infants capable of spatial indexing at such an early age?

Location seems to naturally support decisions of intersensory equivalence in adults. For example, if one sees and hears something coming from a specific location, more often than not one will

assume that the sound and sight are related in a meaningful way, such as thinking that a mannequin is talking (Spence & Driver, 2000). It is not altogether surprising, therefore, that infants can learn to associate coherently co-occurring visual and auditory stimuli that always co-occur in the same location. What is interesting is that they continue to make that association even in the absence of the visual stimulus, and after the location has moved. Perhaps this is an effect of the robust nature of cross-modal information, especially in the service of learning (cf. de Sa & Ballard, 1998). For example, when a spatial cue is presented in only one modality, as is the case in delayed response paradigms, the eye movements of 6-month-old infants reveal that their visuospatial working memory capacity is fragile (Gilmore & Johnson, 1995). In Gilmore and Johnson's (1995) task, 6-month-olds were presented with an abstract visual cue, centrally located, that predicted the location of a visual target. Infants' anticipatory saccades indicated that they maintained spatial representations over time delays of 3–4 s, but that those representations decayed after that. Our paradigm, using a similar dependent measure of eye movements, suggests that these spatial representations can be more robust, and indeed retained for longer, when presented in a cross-modal fashion. In Experiment 3, there was a minimum of 4 s between familiarization and test trial, during which the ports moved. In addition, because there were two different multimodal events during familiarization, there was often one or two intervening 8-s trials between the presentation and test of a particular multimodal event.

A 6-month-old infant, with a fragile visuospatial working memory, who has only just started to use egocentric reference frames, nonetheless looks longer at the location of a past event after a large delay and movement of that location. We argue that the ability to spatially index under these challenging circumstances was supported by the rich, multimodal nature of the events. Developmental researchers studying cross-modal abilities have proposed these intersensory sensitivities to be so important as to be requisite for the development of higher order perceptual and cognitive functions (Lewkowicz, 2000; Thelen & Smith, 1994). As was reviewed in the introduction, infants have a vast repertoire of cross-modal abilities by the age of 6 months. In addition, when infants are presented with stimuli in one modality, they demonstrate rapid learning of the statistical regularities of that input (Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999). Perhaps these two features of infant cognition—intersensory sensitivity and associative learning—have a beneficial interaction. In a parallel case, recent research modeling language learning has shown that multiple probabilistic cues (e.g., lexical stress, phonemes, pauses), although they may be individually unreliable, can be integrated to produce faster learning of such things as word boundaries and syntax (Christiansen, Allen, & Seidenberg, 1998; Christiansen & Dale, 2001; see also Massaro, 1999). Similarly, the multiple, cross-modal cues infants were presented with in our paradigm may have allowed for rapid and robust learning of spatial regularities. Therefore, we suggest there is a developmental importance in picking up on intersensory associations in a dynamic multimodal world.

We have sampled only two points across the life span but have found this spatial indexing ability at both ages. This is an exciting and interesting finding, which opens the doors to more studies investigating the details of the developmental trajectory and the

parameters of the ability. We posit that this type of dynamic spatial indexing is supported by statistical learning and cross-modal sensitivity, both of which are available to very young infants (Kirkham et al., 2002; Lewkowicz, 2000). Would an infant younger than 6 months of age be capable of using the multimodal cues to encode location for memory? Recent research looking at location statistics suggests that by 11 months of age, infants can encode a spatial pattern with only location as the cue (e.g., the same red circle shows up in a statistically probable pattern of locations), but at 8 months of age, infants require redundant shape and/or color information to support learning of a spatial pattern (e.g., they notice the pattern only if the objects in the pattern are distinct colored shapes; Kirkham, Slemmer, & Johnson, 2003). Thus, perhaps the multiple cues in the current paradigm (shape, color, synchronous motion, and sound) could support learning of the location in infants younger than 6 months. However, the statistical learning studies used overall looking time measures to a familiar or novel display, not eye movements to particular regions of the screen. Gilmore and Johnson (1997) have presented compelling evidence that infants younger than 5 months of age represent the targets of their saccades by using retinocentric coordinates. Thus, even if looking time suggests an ability to encode location with enough cues, perhaps it would not be seen in eye movements.

Relation to Adult and Infant Theories of Indexing

There are several paradigms in the literature that have been used to demonstrate forms of spatial indexing, as we discussed earlier. Specific theories of indexing have grown from these paradigms. Do they cohere or clash with the interpretation of our results presented above?

The parallel has already been noted between our task of memorizing factual information (Experiment 1) and the block-moving task used by Ballard et al. (1995). Ballard et al. (1997) modeled performance in this task by implementing a computational theory of deictic coding. Motivated by issues of computational complexity and the physical constraints of the brain, Ballard et al. (1997) claimed that at the time scale of about a third of a second, orienting movements of the body can play a key role in cognitive operations. These orienting movements—usually saccades—are termed *deictic codes*, and it is argued that they can be used to bind external sensory data to internal representations and programs. The claim is that in many natural tasks, working memory computations are vastly simplified by the use of deictic reference.

Ballard et al. (1997) implemented their model of the working memory operations involved in the block-moving task. In this neurally plausible computational framework, deictic codes are supported by location and identification functions. The identification function extracts a set of low-level visual features from the display at the currently fixated location (in the model, this was the central region of a camera's view). The location function searches the whole visual array for a particular set of features and returns the image coordinates of the best-matching region. Combinations of these functions allow the model to rapidly form a pointer, or deictic code, for a particular item on the basis of a low level set of features. When more information about this object is needed, the location function fixates it and the relevant information is extracted. In this way, deictic codes allow for “just-in-time” repre-

sentation, finding, and encoding detailed information only as it is needed in the task.

Ballard et al.'s (1997) model certainly fits the data of adult participants in the block-moving task, but can the theory of deictic codes also be used to describe the behavior of our participants? There are certainly strong commonalities between the just-in-time nature of deictic codes and the concept of external memory discussed above. The comparison breaks down, however, when we consider the subfunctions of deictic codes. In Ballard et al.'s (1997) theory, the referents of pointers are found each time by a location function, which searches the visual field for a certain set of visual features. This function would not be able to discriminate the ports presented in our displays. The two ports are identical in their local features, and because they move between presentation and test phases, they cannot be identified by their global screen positions. Therefore, our studies suggest that elements of the visual array can be continuously tracked, and their locations updated.

Pylyshyn's (1989) theory of visual indexing proposes such a mechanism of object tracking to account for participants' performance in the multiple-object-tracking paradigm (e.g., Pylyshyn & Storm, 1988). Certainly, our participants updated the objects' locations when they moved, in the same way that participants in multiple-object-tracking studies are capable of tracking objects through different locations. There are, however, several important differences that should be noted. Pylyshyn's (1989, 2001) theory has not been tested outside of the multiple-object-tracking paradigm to our knowledge and so cannot be generalized to events that include intersensory information. Furthermore, in that paradigm the explicit goal is to track objects. In our studies, indexing emerges in the eye movements of participants during the task of remembering facts, or simply perceiving multimodal events. Whereas Pylyshyn's work attests to the impressive indexing abilities of adult subjects, our results speak to the way in which indexing spontaneously collaborates with location encoding and multimodal perception.

The Leslie et al. (1998) theory of the infant “object concept” offers converging evidence and arguments for the importance of spatiotemporal information in development. One of their claims is that first spatiotemporal information and, later in development, featural information (e.g., color and shape) can be used to establish an index. Research to date has only considered a narrow set of static, unimodal features when investigating infants' understanding of objects. The degree to which infants associated visual features with a location was not addressed in our paradigm, and therefore, our results cannot speak to the developing role of features in object indexing that Leslie et al. propose. Our results provide good evidence, however, that by 6 months of age infants can associate a sound with a location, track that location as it moves, and then fixate that location when presented with the auditory cue. We suggest, therefore, that by using sensitive measures, such as eye tracking and by presenting moving cross-modal stimuli, researchers can obtain a much richer view of the infant's ability to spatially index.

Conclusion

The perceptual system faces the challenge of a dynamic multimodal environment. This challenge is met in part by the cooperation of location encoding, object tracking, and cross-modal inte-

gration, which allows people to take advantage of the inherent regularities of the natural world as well as important intersensory information. Objects in the world tend to have their own coherent intersensory features (e.g., color, shape, characteristic sounds, predictable motion, motor affordances), all of which co-occur at the same spatial location. This association is so compelling, we argue, that it appears from the earliest point that an infant can reliably represent locations in space, and it continues to drive eye movements of adults even when they are trying to remember spoken facts.

References

- Andrade, J., & Meudell, P. (1993). Short report: Is spatial information encoded automatically in memory? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *46*(A), 365–375.
- Arrington, C., Carr, T., Mayer, A., & Rao, S. (2000). Neural mechanisms of visual attention: Object-based selection of a region in space. *Journal of Cognitive Neuroscience*, *12*, 106–117.
- Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*, 190–201.
- Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, *69*, 1263–1275.
- Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*, 66–80.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723–767.
- Baylis, G., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 451–470.
- Blaser, E., Pylyshyn, Z., & Holcombe, A. (2000, November 9). Tracking an object through feature space. *Nature*, *408*, 196–199.
- Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1988). Perceptually driven movements as contextual retrieval cues. *Bulletin of the Psychonomic Society*, *26*, 541–553.
- Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.
- Brouwer, R. T. F., & Van der Heijden, A. H. C. (1997). Identity and position: Dependence originates from independence. *Acta Psychologica*, *95*, 215–237.
- Calvert, G. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*, 1110–1123.
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, C. K., et al. (1997, April 25). Activation of auditory cortex during silent lipreading. *Science*, *276*, 593–596.
- Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*, 221–268.
- Christiansen, M. H., & Dale, R. A. C. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 220–225). Mahwah, NJ: Erlbaum.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.
- Clark, A. (1997). *Being there: Putting brain, body, and the world together again*. Cambridge, MA: MIT Press.
- Clifton, R. K., Muir, D. W., Ashmead, D. H., & Clarkson, M. G. (1993). Is visually guided reaching in early infancy a myth? *Child Development*, *61*, 1098–1110.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*, 337–367.
- de Sa, V. R., & Ballard, D. H. (1998). Category learning through multimodality sensing. *Neural Computation*, *10*, 1097–1117.
- Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Science*, *2*, 254–262.
- DuBose, C. N., Cardello, A. V., & Maller, O. (1980). Effects of colorants and flavorants on identification, perceived flavor intensity, and hedonic quality of fruit-flavored beverages and cake. *Journal of Food Science*, *45*, 1393–1399.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501–517.
- Egeth, H., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297.
- Egley, R., Driver, J., & Rafal, R. (1994). Shifting visual attention between objects and locations: Evidence for normal and parietal lesion participants. *Journal of Experimental Psychology: General*, *123*, 161–177.
- Ellis, N. (1991). Automatic and effortful processes in memory for spatial location. *Bulletin of the Psychonomic Society*, *299*, 28–30.
- Epstein, R., & Kanwisher, N. (1999). Repetition blindness for locations: Evidence for automatic spatial coding in an RSVP task. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1855–1866.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the visual identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143–149.
- Fenwick, K. D., & Morrongiello, B. A. (1998). Spatial co-location and infants' learning of auditory–visual associations. *Infant Behavior and Development*, *21*, 745–759.
- Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory–visual interactions in human saccadic eye movements. *Perception & Psychophysics*, *57*, 802–816.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.
- Gilmore, R. O., & Johnson, M. H. (1995). Working memory in infancy: Six-month-olds' performance on two versions of the oculomotor delayed response task. *Journal of Experimental Child Psychology*, *59*, 397–418.
- Gilmore, R. O., & Johnson, M. H. (1997). Egocentric action in early infancy: Spatial frames of reference for saccades. *Psychological Science*, *8*, 224–230.
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *6*, 325–331.
- Harris, L. R. (1980). The superior colliculus and movements of the head and eyes in cats. *Journal of Physiology*, *300*, 367–391.
- Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*, 43–64.
- Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, *38*, 125–137.
- Jiang, Y., Olson, I., & Chun, M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 683–702.
- Johnson, M. H., & Gilmore, R. O. (1998). Object-centered attention in 8-month-old infants. *Developmental Science*, *1*, 221–225.
- Johnston, J., & Pashler, H. (1990). Close binding of identity and location in visual feature perception. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 843–856.
- Jousmäki, V., & Hari, R. (1998). Parchment-skin illusion: Sound-biased touch. *Current Biology*, *8*, R190.
- Kahneman, D., & Henik, A. (1981). Perceptual organization and attention.

- In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Erlbaum.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 174–219.
- Kaldy, Z., & Leslie, A. M. (2003). Identification of objects in 9-month-old infants: Integrating “what” and “where” information. *Developmental Science*, *6*, 360–373.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2003). *Location, location, location: Development of visuospatial statistical learning in infancy*. Unpublished manuscript.
- Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*, 513–549.
- Köhler, S., Moscovitch, M., & Melo, B. (2001). Episodic memory for object location versus episodic memory for object identity: Do they rely on distinct encoding processes? *Memory & Cognition*, *29*, 948–959.
- Kuhl, P. K., & Meltzoff, A. N. (1982, February 26). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1140.
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*, 3559–3565.
- Land, M. F., & Lee, D. N. (1994, June 30). Where do we look when we steer. *Nature*, *369*, 742–774.
- Laplante, D. P., Orr, R. R., Neville, K., Vorkapich, L., & Sasso, D. (1996). Discrimination of stimulus rotation by newborns. *Infant Behavior and Development*, *19*, 271–279.
- Leslie, A. M., Xu, F., Tremoulet, P., & Scholl, B. J. (1998). Indexing and the object concept: Developing “what” and “where” systems. *Trends in Cognitive Sciences*, *2*, 10–18.
- Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*, 281–308.
- Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity of matching. *Developmental Psychology*, *16*, 597–607.
- Luck, S., & Vogel, E. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.
- Massaro, D. W. (1999). Speechreading: Illusion or window into pattern recognition. *Trends in Cognitive Science*, *3*, 310–317.
- McDonald, J., & Ward, L. (2000). Involuntary listening aids seeing: Evidence from human electrophysiology. *Psychological Science*, *11*, 167–171.
- McGurk, H., & MacDonald, J. (1976, December 23). Hearing lips and seeing voices. *Nature*, *264*, 746–747.
- Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: A brief overview. *Cognitive Brain Research*, *14*, 31–40.
- Morrongio, B. A., Fenwick, K. D., & Chance, G. (1998). Crossmodal learning in newborn infants: Inferences about properties of auditory–visual events. *Infant Behavior and Development*, *21*, 543–554.
- Most, S., & Alvarez, G. (2002, May). *But it's the only thing there! Sustained inattention blindness for a solitary stimulus*. Paper presented at the annual meeting of the Vision Sciences Society, Sarasota, FL.
- Naveh-Benjamin, M. (1988). Recognition memory of spatial location information: Another failure to support automaticity. *Memory & Cognition*, *16*, 437–445.
- Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, *7*, 480–494.
- Nissen, M. J. (1985). Accessing features and objects: Is location special? In M. Posner & O. Marin (Eds.), *Attention and performance XI: Mechanisms of attention* (pp. 205–220). Hillsdale, NJ: Erlbaum.
- O'Craven, K., Downing, P., & Kanwisher, N. (1999, October 7). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.
- Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception & Psychophysics*, *64*, 1055–1067.
- O'Regan, J. K. (1992). Solving the real mysteries of visual perception: The world as an “outside memory.” *Canadian Journal of Psychology*, *46*, 461–488.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.
- Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial index model. *Cognition*, *32*, 65–97.
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, *80*, 127–158.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277.
- Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*, 269–295.
- Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*, 347–357.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.
- Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science*, *6*, 233–244.
- Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*, 1–46.
- Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple objects through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*, 259–280.
- Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*, 159–177.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000, December 14). What you see is what you hear. *Nature*, *408*, 788.
- Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport*, *12*, 3849–3852.
- Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, *7*, 301–305.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, *28*, 1059–1074.
- Simons, D. J., & Levin, D. (1997). Change blindness. *Trends in Cognitive Science*, *1*, 261–267.
- Slemmer, J. A., & Johnson, S. P. (2002, May). *Object tracking in ecologically valid occlusion events*. Poster session presented at the annual meeting of the Vision Sciences Society, Sarasota, FL.
- Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, *15*, 626–636.
- Spelke, E. S. (1981). The infant's acquisition of knowledge of bimodally specified events. *Journal of Experimental Child Psychology*, *31*, 279–299.
- Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1005–1030.
- Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, *59*, 1–22.
- Spence, C., & Driver, J. (2000). Attracting attention to the illusory location

- of a sound: Reflexive crossmodal orienting and ventriloquism. *Neuroreport*, *11*, 2057–2061.
- Spence, C., Nicholls, M. E. R., Gillespie, N., & Driver, J. (1998). Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision. *Perception & Psychophysics*, *60*, 544–557.
- Spivey, M. J., & Geng, J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*, 235–241.
- Spivey, M. J., Richardson D. C., & Fitneva, S. (in press). Memory outside of the brain: Oculomotor indexes to visual and linguistic information. In F. Ferreira & J. Henderson (Eds.), *Interfacing language, vision, and action*. San Diego, CA: Academic Press.
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*, 73–74.
- Stein, B. E., Magalhaes-Castro, B., & Kruger, L. (1975, July 25). Superior colliculus: Visuotopic–somatopic overlap. *Science*, *189*, 224–226.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
- Stein, J. F. (1992). The representation of egocentric space in the posterior parietal cortex. *Behavioral and Brain Sciences*, *15*, 691–700.
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
- Tipper, S., Driver, J., & Weaver, B. (1991). Object-centered inhibition of return of visual attention. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *43*(A), 289–298.
- Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *40*(A), 201–237.
- Tremoulet, P. D., Lee, N., & Leslie, A. M. (1998, April). *Can 9-month-olds identify by shape?* Poster session presented at the biennial International Conference on Infant Studies, Atlanta, GA.
- Van der Heijden, A. H. C. (1993). The role of position in object selection in vision. *Psychological Research*, *65*, 44–58.
- Von Hofsten, C. (1980). Predictive reaching for moving objects by human infants. *Journal of Experimental Child Psychology*, *30*, 369–382.
- Wheeler, M., & Treisman, A. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48–64.
- Winograd, E., & Church, V. (1988). Role of spatial location in learning face–name associations. *Memory & Cognition*, *16*, 1–7.
- Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, *30*, 111–153.
- Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 458–468.

Received July 2, 2002

Revision received September 26, 2003

Accepted October 23, 2003 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://watson.apa.org/notify/> and you will be notified by e-mail when issues of interest to you become available!