

What's in a look?

Richard N. Aslin

Department of Brain and Cognitive Sciences, University of Rochester, USA

Abstract

The most common behavioral technique used to study infant perception, cognition, language, and social development is some variant of looking time. Since its inception as a reliable method in the late 1950s, a tremendous increase in knowledge about infant competencies has been gained by inferences made from measures of looking time. Here we examine the logic, utility, and future prospects for further gains in our understanding of infant cognition from the use of looking time measures.

Introduction

Consider the following scenario. Three adults enter a crowded room filled with volunteers and paid staff members who are waiting for a political candidate to emerge from behind a curtain and make either a victory or a concession speech. Person A is a security guard assigned to protect the candidate. Person B is the candidate's spouse. Person C is a member of the opposition party. In the interests of science, all three of these people are wearing head-mounted eye-trackers that provide a moment-by-moment record of their patterns of gaze over the next 10 minutes as the candidate approaches the podium and speaks to the audience.

As the candidate appears and walks toward the microphone, persons A, B and C immediately fixate and follow the candidate's face as she prepares to speak, exhibiting no differences in the duration of fixation across the three viewers. Then person A, the security guard, begins to rapidly scan the faces and hands of the entourage surrounding the candidate. Person B, the candidate's husband, continues to look directly at the newly elected governor's face as she explains her landslide victory, only occasionally glancing to the faces of other people on stage who were closely involved in the campaign. Person C, the member of the opposition party, exhibits the same general pattern of fixations as the candidate's spouse.

The point of this example is that, after an initial phase of sustained fixation that was identical among the three people, A and B showed dramatically different patterns of fixation, and B and C showed very similar patterns of fixation. Yet, knowing the relationship of each person to the candidate on stage, we would readily interpret A's

diverse scanning as a search for potential threats rather than disinterest in the candidate (or the presence of an attentional deficiency), B's focal scanning as seeking confirmation of the candidate's joy at winning the election, and C's overtly similar focal scanning as an attempt to glean reactions from the candidate that might be used strategically in future elections to return the governorship to the opposition party. In sum, differences in looking times typically reflect different underlying cognitive processes, but the same duration of looking may also reflect quite different underlying cognitive processes (e.g. orienting, search for discrepancy, sustained attention with positive affect, sustained attention with null or negative affect).

Since the pioneering studies of looking behavior in young infants by Robert Fantz in the late 1950s, there have been dozens of fixation paradigms developed to test various aspects of infants' detection, discrimination, preference, categorization, learning, and expectations of both visual and auditory stimuli. It is no exaggeration to say that without looking time measures, we would know very little about nearly any aspect of infant development. In the early 1960s, mainstream ophthalmology textbooks claimed that newborns were blind, audiology textbooks claimed that newborns were deaf, and researchers who relied on more sophisticated motor responses like reaching claimed that cognitive abilities were extremely rudimentary until the end of the first postnatal year. But despite tremendous advances over the past 50 years in what we have learned about infant development, caution must be exercised in what conclusions are drawn from looking time data. Like the introductory example, duration of looking to a stimulus, in comparison to some 'control' stimulus, serves as the dependent measure of some putative (i.e. experimenter defined) cognitive

Address for correspondence: Richard N. Aslin, Department of Brain and Cognitive Sciences, Meliora Hall, River Campus, University of Rochester, Rochester, NY 14627, USA; e-mail: aslin@cvs.rochester.edu

state. Clearly, this is a many-to-one mapping problem: many potential 'hidden' variables contribute to a single dependent measure. How then do we make sense of looking time data?

Operational definition of a look

There are, of course, many different metrics one could use to characterize looking behavior in adults or infants, including total looking time, average fixation duration, time to habituation, direction of first look, frequency of switching between simultaneously visible objects, etc. The key to interpreting any given metric is the linking hypothesis that joins the dependent variable to the underlying cognitive process. What, then, are the possible linking hypotheses for looking times (see Teller, 1984)? The answer is determined, in large part, by the question that is the focus of the specific research paradigm. For example, simultaneous two-alternative forced-choice (2AFC) paradigms are used to assess stimulus detection (compared to a no-stimulus control). In this case the linking hypothesis is straightforward: percent correct is presumed to map monotonically onto increasing stimulus visibility.

In contrast to detection, discrimination paradigms come in so many flavors that the linking hypothesis is often undefined and typically unclear. Consider the simplest possible extension of the 2AFC detection paradigm that replaces the no-stimulus control with a suprathreshold comparison stimulus. If the percent preference for one stimulus (the target) is significantly greater than chance (50%), then discrimination can be inferred. But what is the relationship between above-chance preference and discrimination? Here the linking hypothesis breaks down, resulting in yes/no conclusions about discrimination rather than the more graded visibility that is inferred from the percent correct output of detection paradigms (see Aslin & Fiser, 2005).

Another problem is that the absence of preference does not imply the absence of discrimination; infants may prefer neither of two stimuli, yet be quite capable of discriminating them from each other. To address this problem, paradigms have been developed to motivate infants to change their intrinsic preferences (or absence thereof) by preceding the test trials with one or more biasing stimuli. Familiarization and habituation paradigms expose infants to one stimulus (or class of stimuli) to induce a decrement in the salience, discriminability, or preference for that stimulus. This decrement can only occur if some underlying mechanism in the infant's brain retains information about the characteristics of the repeating stimulus (i.e. a form of visual memory). Thus,

stimulus discrimination and visual memory are confounded in any familiarization or habituation paradigm (e.g. side-by-side discrimination is often superior to successive discrimination; see Oakes & Ribar, 2005). Nevertheless, significant recovery of looking time, compared to a no-change control, during post-familiarization (or post-habituation) test trials forces the conclusion that the repeated stimulus was discriminated from the novel (non-repeated) stimulus.

The discovery that a biasing stimulus could alter looking times on post-familiarization (or post-habituation) test trials raised the question of how much biasing experience is sufficient to induce a preference when none existed prior to familiarization or habituation. Different paradigms have settled on different criteria, largely through trial and error. Familiarization paradigms tend to use fixed trial durations (or even a single familiarization trial), whereas habituation paradigms tend to use the 'industry standard' of a 50% decline in mean looking time across three successive trials compared to the initial three trials. Habituation paradigms also tend to use trial durations determined by the infant rather than by the clock. A look away from the stimulus that exceeds 1 or 2 sec serves to terminate that trial. In contrast, familiarization paradigms tend to have a preset number of exposure trials or a preset cumulative looking time across an individually determined number of exposure trials.

The problem with these various familiarization or habituation criteria for inducing a bias in test-trial looking times is that there is no anchoring of the obtained results as there is in the 2AFC detection paradigm. That is, when a stimulus is paired with a no-stimulus control and the stimulus is varied across test trials along some physical dimension (e.g. luminance, contrast, stripe width), the performance (percent correct) of the infant varies between chance (50%) and some asymptotic level (e.g. 90%). In biasing paradigms, however, the infant's test performance, as in any preference paradigm, is either significantly above chance (50%) or not; there is no graded linking hypothesis between percent preference and discriminability. As a result, only positive evidence of discrimination appears in the published literature; negative results suffer from the problem of failing to refute the null hypothesis. Comparisons across age are also problematic with biasing paradigms because of the confound between visual memory and stimulus discrimination. Without a function that maps a dependent measure (like percent correct) to discriminability, results at a given age could be attributed to visual memory, independent of discriminability.

A final problem with preference and biasing paradigms is that they rely on a global measure of looking time. Typically, one or at most two stimuli are presented repeatedly and the duration of looking per trial (minus

look-aways less than 2 sec) is the sole dependent measure. Other potential metrics, such as the sequence and duration of fixations that fall within the look-away criterion, the number of look-aways, and the minimum and maximum look durations, are discarded. That is, the field has settled on measures of the *macrostructure* of looking times and ignores the *microstructure*. This is in large part the result of the technical demands (and expense) of automated eye-trackers, which have been commercially available since the early 1980s but used by only a few infant labs (see special Thematic Collection in *Infancy* with introduction by Aslin & McMurray, 2004). One implication of the focus on the macrostructure of looking times is that all sub-looking time metrics are treated as error variance. That is, not only are looking times composed of individual fixations of variable duration, but all looking times include a mix of active information processing and blank stares. In the absence of a more direct measure of neural information processing, a look is merely a correlate of the underlying neural activity that mediates detection, discrimination, or categorization of visual stimuli in a specific behavioral paradigm.

Paradigm complexity and linking hypotheses

The use of looking times to infer underlying cognitive processes expanded its territory beyond detection and discrimination in the late 1970s with studies of category formation. The key insight was to create a biasing paradigm with *variable* exemplars rather than a single familiarization (or habituation) stimulus. If infants formed a category (or common dimension) during this biasing phase, then their performance on the test phase should lead them to recover looking time to exemplars from a novel category, but not to recover to exemplars from the familiar category. Of course, there is no reason why infants should fail to discriminate within-category differences since adults (and non-humans) are certainly capable of doing so under the right testing conditions. Thus, to be definitive, the pattern of results had to meet a high (and unrealistic) standard (significant between-category recovery and no within-category recovery). This criterion was subsequently softened to include any evidence of greater between- than within-category recovery. Such a modified pattern of results was interpreted as evidence that the underlying commonality that defined the category had been extracted during the biasing phase.

The problem with the foregoing logic of multiple-exemplar biasing paradigms is that, depending on the relative salience of stimulus differences that define the category of interest compared to stimulus differences that

are not definitional (i.e. irrelevant) to the category, any pattern of results can be obtained. For example, small within-category differences guarantee strong between-category recovery of looking times, whereas large within-category differences will certainly mitigate the magnitude of recovery. Without an independent estimate of what qualifies as 'small' or 'large', the only way to predict the pattern of results is to collect the data with a given set of stimulus materials and hope that the outcome of the experiment fits a canonical pattern of looking times. Given the vagaries of stimulus salience (however defined) and the likely change in these patterns of salience at different infant ages, it is not surprising that such designs are relatively rare in the literature.

An alternative design is to abandon the multiple-exemplar biasing phase and replace it with a single exemplar, but one that contains at least two underlying dimensions of interest. The logic of this paradigm is that if infants preferentially extract (or attend to) information in one of the two dimensions, then test trials that isolate that dimension will be looked at less than test trials depicting the non-extracted dimension. While this paradigm works well in many cases, it also suffers from the problem of dimensional salience: by making the 'relevant' dimension more salient one can shift the pattern of results toward sensitivity to that dimension and away from sensitivity to the 'irrelevant' dimension. As with the multiple-exemplar design, dimensional salience is likely to change with infant age (and experience).

A final design alternative is to abandon the biasing phase entirely and presume that test trials assess whatever preferences the infant brings to the lab from innate biases or from prior learning via experience in the natural environment. Such a paradigm, of course, asks a different question than biasing paradigms. The latter assesses whether intrinsic biases can be altered, whereas the former assesses the biases accumulated with no immediately prior exposure. As one might expect, only the most robust preferences that infants bring to the lab are expressed when there is no immediately preceding familiarization (or habituation). This brings us full circle to the fundamental limitation of basic preference paradigms: without an extrinsic down-weighting of stimulus salience by repetition, infants may show no spontaneous preference for one of two stimuli despite the ability to discriminate between them.

The foregoing paradigms, all of which attempt to draw inferences about cognitive processes that go well beyond simple detection or discrimination, suffer from a lack of clarity about the underlying linking hypothesis. This lack of clarity is exemplified by the diversity of terms used to describe the infant's pattern of looking times on the test trials: surprise, reasoning, category

formation, object completion, number sense, core knowledge, etc. A single dependent measure cannot definitively reflect all of these underlying constructs unless there is some way to partition looking time into its components. But if only the macrostructure of looking times is being assessed, then all we know is that *some* aspect of the relationship between what the infant brings to the lab and the exposure they receive during the experiment is contributing to their behavior during the test phase. I have argued elsewhere (Aslin, 2000) that the systematicity of results across a large number of experiments, using the same paradigm and subtle variations in stimuli, can converge on a sensible interpretation of an overall pattern of results. But care must be taken to limit the linking hypothesis from looking times to underlying constructs in the most conservative way possible, including using terminology that posits the simplest underlying mechanism.

This emphasis on the pattern of results across many experiments must confront another problem for the field of infant cognition: the presence of post-biasing familiarity effects. Since the advent of familiarization and habituation paradigms in the 1970s, it has been observed in some studies that infants do not look longer at the more novel test stimulus. These rare findings fly in the face of the linking hypothesis that posits a decline in attention/salience with stimulus repetition and a concomitant growth in the encoding/memory of that stimulus. Because these familiarity effects can occur even after infants have met the standard criterion of habituation (see Fiser & Aslin, 2002), one cannot attribute them to inadequate familiarization. Two possible conclusions are forced by these findings. One is that infants presented with a complex or highly diverse set of stimuli during the habituation phase can meet the 50% decrement criterion without having 'fully encoded' the stimuli. As suggested by several models of looking time (Hunter & Ames, 1988; Rose, Gottfried, Melloy-Carminar & Bridger, 1982; Wagner & Sakovits, 1986), less than full encoding (e.g. during the initial phase of familiarization) leads infants to seek familiar stimuli over novel stimuli (see also Roder, Bushnell & Sasseville, 2000). The other interpretation is that sub-components (or dimensions) of 'fully encoded' stimuli can trigger a *recognition* response for familiarity that is stronger than the *attentional* response to novelty (see Kaplan & Werner, 1986, for a similar two-process model). These two alternatives are very similar in positing a 'balance' between underlying cognitive processes that compete to yield a single outcome: looking time on test trials. Without an independent assessment of these confounded cognitive processes, however, it will not be possible to predict whether familiarity or novelty effects will be obtained in a given experiment.

A final dilemma for researchers who use looking times to make inferences about the processing of *auditory* stimuli is the role of cross-modal competition. The beauty of looking times is that they are elicited spontaneously from infants at all ages; they do not require a training phase like high-amplitude sucking or visual reinforcers like conditioned head-turning. The discovery by Horowitz (1975) that duration of looking to a neutral *visual* stimulus can also serve as a spontaneous measure of auditory processing created an explosion of research on infant speech and language development (e.g. Best, McRoberts & Sithole, 1988). But the paradigms that emerged created a cross-modal conflict that was not present in looking time paradigms used in the visual modality. When an infant looks at a visual stimulus, the linking hypothesis is that gaze serves a useful purpose for visual processing. But when an infant looks at a visual stimulus while listening to an auditory stimulus, the linking hypothesis is less direct. Presumably, auditory processing that is ongoing during fixation of the visual stimulus creates a delay in how likely the infant's visual system will disengage attention from that visual stimulus. Although it is natural for most organisms to look in the direction of multimodal objects (i.e. objects that have both visual and auditory properties), the neural mechanisms that mediate looking time behaviors to auditory stimuli are certainly more complex than those that mediate looking time behaviors to visual stimuli. This accounts, in part, for the fact that researchers struggle to find the 'optimal' visual stimulus at a given infant age that enables looking times to show a range of variability with different auditory stimuli (i.e. neither floor nor ceiling effects). But of course this struggle highlights the difficulty of drawing cross-age conclusions even if the same visual stimulus is employed.

A look to the future: converging measures and predictive models

Both the novice and the expert in the field of infant development could easily interpret the foregoing discussion of methodological dilemmas as motivation to seek employment in another scientific discipline. I am much less pessimistic because there are currently available paradigms and techniques that can, if used judiciously, remedy most of the problems of interpretation.

First, it is well within the capacity of the field to gather data from separate groups of infants (at different ages) to establish intrinsic preferences or measures of stimulus salience. When such data are available, they have dramatic effects on discrimination performance using looking time paradigms (Civan, Teller & Palmer, 2004; Kaldy, Blaser & Leslie, 2006). It is difficult to gather such data,

but as the field asks more subtle questions, it will become essential.

Second, the typical experimental design counterbalances all 'nuisance' variables (e.g. stimulus order during familiarization and test phases, non-criterial stimulus dimensions, etc.). However, such counterbalancing presumes that the effect of interest is sufficiently robust that it can survive all of the variability introduced by the sequence of different events presented to a given infant. For example, while in principle a diversity of test trial types can be presented to each infant, this diversity can induce contrast effects that would not be present if only a single type of test trial were presented. An alternative design that gathers only a single data point from each participant requires many more infants to be tested, but it has proven essential in studies of the spontaneous behavior of non-human primates and should be used more commonly in studies of human infants (see Feigenson, Carey & Hauser, 2002).

Third, converging measures allow for the simultaneous collection of behavioral and physiological data that could clarify a global measure like looking time. Such combined measures are already being used in some labs to gather from individual infants, simultaneously, both looking times and other dependent measures such as heart rate (Colombo, Richman, Shaddy, Follmer Greenhoot & Maikranz, 2001; Richards, 1987) and ERPs (Reynolds & Richards, 2005). This enables the investigator to use two measures, each with their own linking hypothesis, to narrow the range of interpretations that are possible given complementary constraints.

Fourth, eye-tracking techniques are now sufficiently easy to use with infants that an examination of the *microstructure* of looking times will become more readily available to a large number of infant labs (von Hofsten, Dahlstrom & Fredriksson, 2005). These techniques have the potential to reveal a wealth of new information about the relationship between global looking times and their subcomponents, such as the number and duration of fixations to elements within a complex stimulus. For example, one reason why infants may show familiarity preferences, or fail to show novelty preferences, is that they have not looked at the critical sources of information in the displays that enable encoding of visual memories of that information (Johnson, Slemmer & Amso, 2004).

Summary and conclusion

In this essay, the complexities of gathering looking time data from human infants, and drawing inferences about underlying (hidden) cognitive processes, was reviewed. Like any discipline, there are many challenges associated

with gaining access to data that are as close as possible to the hidden constructs that could clarify a specific underlying mechanism. Looking times have been a very productive measure over the past 40 years of research on infant sensation, perception, cognition, and language development. However, as the research questions have become more subtle, current methods must be enhanced and new methods must be developed (alone or in combination with old methods). In addition, care must be taken to be explicit about the linking hypotheses that are used to draw inferences about underlying cognitive processes. The future of the field will be brighter if baseline estimates of key parameters are used to make model-predictions that are then tested empirically, and more fine-grained techniques such as eye-tracking are used to uncover the microstructure of global looking time measures.

Acknowledgement

Preparation of this essay was enabled, in part, by a program grant from the J.S. McDonnell Foundation (21002089).

References

- Aslin, R.N. (2000). Why take the cog out of infant cognition? *Infancy*, *1*, 463–470.
- Aslin, R.N., & Fiser, J. (2005). Methodological challenges for understanding cognitive development in infants. *Trends in Cognitive Sciences*, *9*, 92–98.
- Aslin, R.N., & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy: methodological developments and applications to cognition. *Infancy*, *6*, 155–163.
- Best, C.T., McRoberts, G.W., & Sithole, N.M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 345–360.
- Civan, A., Teller, D.Y., & Palmer, J. (2004). Relations between spontaneous preferences, familiarized preferences, and novelty effects: measurements with forced-choice techniques. *Infancy*, *7*, 111–142.
- Colombo, J., Richman, W.A., Shaddy, D.J., Follmer Greenhoot, A., & Maikranz, J.M. (2001). Heart-rate defined phases of attention, look duration, and infant performance in the paired-comparison paradigm. *Child Development*, *72*, 1605–1616.
- Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: object-files versus analog magnitudes. *Psychological Science*, *13*, 150–156.
- Fiser, J., & Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*, 15822–15826.
- Horowitz, F.D. (1975). Visual attention, auditory stimulation, and language discrimination in infants. *Monographs of the*

- Society for Research in Child Development*, **39** (5–6), Serial No. 158.
- Hunter, M.A., & Ames, E.W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In L.P. Lipsitt (Ed.), *Advances in child development and behavior* (pp. 69–95). New York: Academic Press.
- Johnson, S.P., Slemmer, J.A., & Amso, D. (2004). Where infants look determines how they see: eye movements and object perception performance in 3-month-olds. *Infancy*, **6**, 185–201.
- Kaldy, Z., Blaser, E., & Leslie, A.M. (2006). A new method for calibrating perceptual salience across dimensions in infants: the case of color vs. luminance. *Developmental Science*, **9**, 482–489.
- Kaplan, P.S., & Werner, J. (1987). Habituation, response to novelty, and dishabituation in human infants: tests of a dual-process theory of visual attention. *Journal of Experimental Child Psychology*, **14**, 83–109.
- Oakes, L.M., & Ribar, R.J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy*, **7**, 85–98.
- Reynolds, G.D., & Richards, J.E. (2005). Familiarization, attention, and recognition memory in infancy: an ERP and cortical source localization study. *Developmental Psychology*, **41**, 598–615.
- Richards, J.E. (1987). Infant visual sustained attention and respiratory sinus arrhythmia. *Child Development*, **58**, 488–496.
- Roder, B.J., Bushnell, E.W., & Sasseville, A.M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy*, **1**, 491–507.
- Rose, S.A., Gottfried, A.W., Melloy-Carminar, P., & Bridger, W.H. (1982). Familiarity and novelty preferences in infant recognition memory: implications for information processing. *Developmental Psychology*, **18**, 704–713.
- Teller, D.Y. (1984). Linking propositions. *Vision Research*, **24**, 1233–1246.
- von Hofsten, C., Dahlstrom, E., & Fredriksson, Y. (2005). 12-month-old infants' perception of attention direction in static video images. *Infancy*, **8**, 217–231.
- Wagner, S.H., & Sakovits, L.J. (1986). A process analysis of infant visual and cross-modal recognition memory. In L. Lipsitt & C. Rovee-Collier (Eds.), *Advances in infancy research* (Vol. 4, pp. 195–217). Norwood, NJ: Ablex.