The Official Journal of the International Society on Infant Studies
Infancy

Infancy, ***17***(1), 126–140, 2012
Copyright © International Society on Infant Studies (ISIS)
ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/j.1532-7078.2011.00097.x

WILEY-BLACKWELL

Infant Eyes: A Window on Cognitive Development

Richard N. Aslin
Department of Brain and Cognitive Sciences
University of Rochester

Eye-trackers suitable for use with infants are now marketed by several commercial vendors. As eye-trackers become more prevalent in infancy research, there is the potential for users to be unaware of dangers lurking "under the hood" if they assume the eye-tracker introduces no errors in measuring infants' gaze. Moreover, the influx of voluminous data sets from eye-trackers requires users to think hard about what they are measuring and what these measures mean for making inferences about underlying cognitive processes. The present commentary highlights these concerns, both technical and interpretive, and reviews the five articles that comprise this Special Issue.

INFANT EYES: A WINDOW ON COGNITIVE DEVELOPMENT

One of the most gratifying moments in the early phase of parenthood occurs when your 6-week-old looks directly at your face—not at the top of your head or slightly to the right or left, but at your eyes, the *real you*. We are exquisitely sensitive to the direction of an infant's gaze (as we are for the gaze of adults), and we make interpretations about psychological processes going on inside the infant's brain based on these gaze patterns. When an infant looks longer at one stimulus over another, whether presented simultaneously or successively, we call that *preference*. When an infant looks away

Correspondence should be sent to Richard N. Aslin, Department of Brain and Cognitive Sciences, Meliora Hall, River Campus, University of Rochester, Rochester, NY 14627. E-mail: aslin@cvs.rochester.edu

from a caregiver who failed to soothe them when they became upset, we call that *avoidance*. When an infant looks longer at an event that, to an adult, is impossible, we call that *violation of expectancy*.

As noted in an earlier review on this topic (Aslin, 2007), I made the following rather provocative statement: "It is no exaggeration to say that without looking time measures, we would know very little about nearly any aspect of infant development" (p. 48). In that review, I also made the point that the vast majority of the data on infant looking have been gathered by observers who make online judgments about whether a stimulus display is fixated or not (a binary metric). In recent years, with the advent of easier-to-use eye-trackers, there has been a shift from studies of the *macrostructure* of looking behavior to the *microstructure* of patterns of fixation (a graded metric). The focus of the present commentary is on the pros and cons of examining the microstructure of infant gaze (and pupil size). What have we learned and, perhaps more importantly, what *could* we learn under ideal conditions using these measures?

I begin with a sobering thought raised in a philosophy of science course that I took as a graduate student (circa 1973). The professor, Keith Gunderson, posed the following question: What if one had access to the firing patterns of every neuron in the brain? Would you know more about how the brain works, and would this knowledge provide useful insights into how the mind works? Setting aside the daunting task of analyzing such a large database of information, Gunderson's query pointed out that the availability of detailed information about a complex system (like the brain) is a blessing and a curse. It is only as beneficial as the theory that links the data to underlying explanatory constructs. For example, what aspect of neural responses is most relevant to understanding *cognition*—is it the firing rates, the timing of neural spike trains, the number of neurons involved in a computation, the connectivity among brain regions, or all of the above?

Gunderson's query is no less relevant to studies of the microstructure of gaze patterns. We know from Yarbus's (1967) classic study of adults that when looking at a complex scene, the pattern of gaze varies depending on the instructions given to the participant. But what exactly are the features of those gaze patterns that matter for *cognition*—is it the duration of each fixation, the spatial spread of clusters of fixations, the back and forth of fixations among key visual elements, or all of the above? And what about the possibility that infants look without seeing? That is, at least some of the time they are surely engaged in a *blank stare*. The point of these introductory comments is to remind infancy researchers that if a new technology enables them to easily collect more data, it is not clear that more is better. We have 50 years of experience with studies that rely on the macrostructure of infants' looking behavior. Although detailed

measurements of infant gaze using photographic techniques were pioneered in the 1930s (McGinnis, 1930) and perfected in the 1960s (Salapatek & Kessen, 1966), there has been an explosion of research using eye-trackers in the past decade. Nevertheless, the old adage is no less relevant today: *Be careful what you wish for!*

DATA QUALITY

The most obvious advantage of an eye-tracker over a human observer is the improvement in resolution, both spatial and temporal. Rather than a binary look-at versus look-away judgment made every 400–500 msec, typical eye-trackers gather samples every 8–20 msec with an accuracy of 1–2° (i.e., 100 or more subregions of a 20 × 20° display). Despite this seemingly beneficial improvement in resolution, it is common for infancy researchers to elicit disdainful expressions from adult and animal researchers when these resolution figures are described. That is because cooperative adults and highly trained animals can provide eye-tracking data with temporal resolutions of 1 msec and spatial resolutions of 1/60th of a degree. And these resolutions enable researchers to ask more sophisticated questions about adults and animals. For example, it is possible to ask on which word in a page of text an adult is fixating, and it is possible to update that text in real-time *during* a shift in fixation (i.e., a 30-msec saccade). These are questions that are simply beyond the capability of any current, and likely any future, eye-tracking system with infants.

What are the limits on temporal and spatial resolution in infant eye-tracking? Temporal resolution is essentially unlimited. In fact, electrooculography (EOG; placing small electrodes on the surface of the face next to each eye) provides an analog signal that can be sampled at any rate, and several EOG studies of infants have collected samples every 5 msec (e.g., von Hofsten & Rosander, 1997). If one wants to document the velocity or acceleration of the eye or what an infant can do on a *single* trial, such high sampling rates are essential. But in most studies of infant cognition, data collected from infants are averaged across trials and across participants. This averaging has the effect of trading off within- and between-subject variance, thereby rendering high temporal resolution unnecessary for most studies of infant cognition. An exception are study designs that rely on *anticipatory* eye movements—if the sample rate is slow relative to the duration of the anticipatory period when a response must be measured, then the magnitude of the anticipatory effect may be subthreshold. Here, the *validity* of timing is crucial—the synchronization between movements of the eye and the presentation of events in the stimulus display. There are several

fundamental ways in which these timing relationships can go awry (see Appendix for details on these timing errors and how to measure them).

Spatial resolution is limited by the type of sensing system and the quality of the calibration data. All video-based eye-trackers have a spatial resolution that is limited by the relationship between the number of pixels in the CCD camera and the size of the optical image of key features of the eye (e.g., the pupil and iris). If the camera has 800 pixels along the horizontal dimension and the pupil or iris border moves 5 pixels/degree, then it is the ability of the hardware or software to reliably detect a 1-pixel shift that limits the resolution of the system (in this case to 1/5th of a degree). But even with an eye-tracker that has such a *system resolution* of 1/5th degree, how the eye-tracker maps these *changes* in estimated gaze position onto the actual display that the participant is viewing is entirely dependent on the accuracy of the calibration. Each participant must fixate known locations on the display (typically five to nine such locations), while the eye-tracker collects baseline data. If the participant is inaccurate in how they fixate these calibration locations, then no matter how good the resolution of the system, the *accuracy* of the resultant gaze estimates will be no better than the accuracy of the calibration. These calibration inaccuracies can result from *variability* in how the participant directs their gaze to each calibration target and from errors in how the software attempts to provide a "best fit" of these calibration data to the known locations of the targets on the display. This best fit is a compromise—it is not feasible to gather calibration data from every location on the display, and so simplifying assumptions are made to take "distorted" calibration data from five to nine locations and force them into a rectilinear grid on the display, thereby allowing any location on the display to have an interpolated x, y value.

Once calibration data have been collected, a key question is what aspect of the gaze data is relevant to a given research question. At one extreme, the eye-tracker could simply be used as an automated observer, rendering judgments of on-screen versus off-screen looking. An intermediate level is to employ algorithms built into the eye-tracker software to parse the gaze pattern into *fixations*. At the other extreme, one could ignore the fact that there are fixations and eye movements, especially because eye movements come in various types that are difficult, even in adults, to categorize (e.g., saccades, smooth pursuit, slow drifts) and simply define areas of interest (AOIs). Any x, y sample from the eye-tracker that falls within an AOI is considered relevant (i.e., an estimate of attention or information processing), even though a small fraction of these x, y samples will consist of a saccade that "flew over" the AOI and happened to be captured during this flight-path. As adults do not process information during saccades, this small proportion of "moving"

x, *y* samples would only slightly inflate any estimates of attention to the stimulus in the AOI.

The foregoing summary of temporal and spatial resolution leads us to several conclusions about using eye-trackers with infants. First, temporal resolution is only relevant for time-critical gaze behaviors, such as visual anticipations (McMurray & Aslin, 2004) or online processing of language (Fernald, Swingley, & Pinto, 2001). Although such behaviors are crucial for some domains of infant cognition, in many domains they are not, and therefore, the 50- to 120-Hz sampling rates of infant eye-trackers are often quite adequate. Second, spatial resolution is not primarily limited by the eye-tracker, but rather by the quality of the calibration (a combination of how well the calibration targets are fixated and how well the software maps these data onto the display screen). Small flashing (or shrinking) targets work well with infants, but the accuracy is unlikely to ever be better than 1° because infants (like naïve adults) do not fixate small stimuli precisely and reliably with the same part of the retina. If 1° of resolution is insufficient to answer a particular question (e.g., whether infants fixate the right nostril or the left nostril of the mother's nose), then an eye-tracker should not be used and alternative methods should be implemented.

QUESTIONS ASKED AND (SOMETIMES) ANSWERED

Given access to an eye-tracker, infancy researchers are like the proverbial hammer in search of a nail. There is a tendency to gather data first and ask questions later. This is a bad idea for three reasons: (1) the data may be of poor quality, (2) there is so much data that something significant will fall out of the analyses by chance, and (3) to verify exploratory (correlational) research you need to follow-up with experiments that test specific hypotheses. Each of these concerns was touched on by the five articles that comprise the present Special Issue.

Morgante, Zolfaghari, and Johnson (in press) addressed the issue of data quality directly by comparing infants and adults on a simple calibration task using one particular eye-tracking system (Tobii T60XL running Studio software; Tobii Technology, Falls Church, VA). There are many potential components of an eye-tracking system that could account for errors in timing or spatial accuracy. First, there is the hardware and "hidden" software (called firmware) that is not accessible to the user. The system provides *x*, *y* coordinates (and pupil size) every 33 msec. Second, these "raw" data must be calibrated so that the *x*, *y* coordinates map onto the stimulus display with high reliability and validity. As noted earlier, the ability of the participant to reliably fixate the calibration targets sets an upper limit on the accuracy of

spatial estimates of gaze position. But the quality of these spatial estimates is also determined by the interpolation algorithms used to map x, y coordinates from the five to nine calibration locations onto the entire display screen. Third, the Studio software collects these gaze estimates and assembles them into a "playback".avi file so that the experimenter can see how the position of gaze is mapped onto the stimulus viewed by the infant. Here is where further errors can be introduced, especially if the stimulus display is dynamic. If the display consists of a moving ball and the question is whether the infant's gaze is reliably tracking (fixated on) the ball, then any software error in how the timing of the x, y gaze estimates and the timing of the changes in the ball's position on the display screen are synchronized will provide faulty estimates of the infant's tracking accuracy. Fourth, there can be timing errors in how the software (e.g., E-Prime [Psychology Software Tools, Inc., Sharpsburg, PA], PsyScope [Freeware, Barcelona, Spain], or Matlab [MathWorks, Natick, MA]) that creates and displays the stimuli on the viewing screen and the datastream of x, y coordinates from the eye-tracker is synchronized. When the display is controlled by one computer and the data collection by another computer, these synchronization issues can be complex, resulting in both constant errors (e.g., a fixed delay) and variable errors (e.g., a drift in the delay across trials).

Given all these potential errors in both the spatial position of estimated gaze position and the timing relationship between gaze position and stimulus position, it is not surprising that most users simply trust that the manufacturer knows what they are doing and that you can expect accurate and reliable data "out of the box." Unfortunately, as noted by Morgante et al. (in press), this is not always a justifiable assumption. As noted by Oakes (2010), it is incumbent on each laboratory to verify the accuracy of their eye-tracker as part of the setup for each experiment, especially because errors can come and go with even minor changes in hardware (e.g., CPU memory and video cards) and software (e.g., system-level settings and latest updates from the manufacturer). Although Morgante et al. report some fairly serious drifts in the accuracy of calibration over trials, the overall level of accuracy is quite good (given all the foregoing caveats). The fact that infants and naïve adults achieve approximately equal calibration errors of slightly over 1° is remarkable, unless of course the design of a specific experiment requires higher levels of accuracy (in which case a different eye-tracker should be used). It is also important to point out that there are software packages (e.g., Shukla, Wen, White, & Aslin, 2011) that bypass all of a given manufacturer's software, thereby ensuring that spatial and temporal errors are minimized. And as noted earlier, there are definitive ways to measure timing errors (see Appendix).

Assuming that data quality is good (or good enough to address a particular research question), then the second warning about using an eye-tracker is

the voluminous data set that is provided. Because eye movements were first recorded from adults over a century ago (Dodge, 1907), you would think that by now we would have a clear idea about which parameters of eye movements and fixations matter most for cognition. Sadly, we do not, except for some crude metrics (e.g., where you look matters because peripheral vision is quite poor). The situation with infants is even less clear, although we do know that there are some rather dramatic developmental changes in oculomotor control (e.g., smooth pursuit tracking is quite poor until several weeks after birth). Sometimes these developmental changes are obvious to the experimenter's "eye"; that is, one can casually glance at eye-movement data and developmental differences are readily apparent (e.g., saccades in young infants consist of repetitive small jumps). But given a large collection of eye-movement data, how does one begin to parse it into relevant and irrelevant parameters that could, in principle, be related to underlying cognitive processes?

Yu, Yurovsky, and Xu (in press) provide an interesting answer to this dilemma by capitalizing on the "eye" of the experimenter to look at the raw (or almost raw) data and to "see" any obvious (and often unexpected) effects. This is a version of the naturalistic observation approach to studying human behavior, except that the infant viewing a stimulus and having their eye movements recorded is far from being in "the wild." There is no question that Yu et al.'s approach has merit—it cuts through the mass of data and looks for the low hanging fruit. But this approach also has some potential pitfalls. First, although there may be what seems like a huge amount of data, it is not clear that this data set contains a large number of "relevant" events that bear on a given aspect of cognition. For example, if infants make two saccades per second and are awake 10 hr per day, then they make 72,000 saccades per day. This sounds like a big number, but how many of these eye movements are used to explore a particular stimulus such as the mother's face? And of those face-directed fixations, how many actually encode some feature of the face? Thus, there is a potential problem of "cognitive sparsity" despite a sea of data. Unfortunately, with sparse data comes the likelihood of false correlations—when two things seem to go together (e.g., looking at the mother's eyes and smiling), the seductive conclusion is that gazing at the mother's eyes triggered the infant's smile. But there were hundreds of other variables that were not measured (e.g., the state of the infant's intestinal system) that could have played a causal (or partially causal) role.

Moreover, the "eye" of the observer will likely miss some important effects because they are small in magnitude. But just because the size of an effect is small does not necessarily imply that the *effect-size* is small. One can have a small effect with small variance that provides more reliable

information than a large effect with even larger variance. This could lead to a problem with the proposed data-analytic strategy proposed by Yu et al. (in press). If the experimenter looks for obvious patterns in the data and then defines an AOI as the focus of a follow-up experiment, then one might well find that the AOI serves as a predictor of some cognitively relevant outcome. But this is like the man who lost his keys and focuses his search only under the street lamp—it may or may not yield useful insights if the street lamp (the AOI) is ill-chosen or the keys (the proportion of fixations) are not the relevant eye-movement parameter. Fortunately, Yu et al. propose a number of protections against such misleading exploratory paths. Two additional conservative strategies they did not mention are (1) the use of cross-validation techniques to iteratively withhold some of the data as a test of hypotheses that were discovered with the remaining data and (2) hierarchical regression techniques to determine which of the myriad of coded variables matter (and by how much). Finally, it is incumbent on data-mining approaches to always follow up with specific tests of hypotheses that have been suggested by preliminary analyses.

Given good-quality data and a variety of compelling hypotheses based on data-mining techniques, how are these hypotheses actually implemented in an experiment with infants using an eye-tracker? One obvious response is to present visual stimuli on a display screen and simply record the infant's eye movements. But this paradigm raises a concern about "naturalness"—the everyday environment consists of real three-dimensional (3-D) objects and surfaces, as well as people and other animate objects. Although such an environment can be simulated on a 2-D video display, what cannot be simulated are the infant's own movements and interactions with that environment. Thus, a spatially fixed eye-tracker display is a technical convenience that enables eye-movement data to be collected with optimal quality. A legitimate worry is that such a constrained "passive" eye-tracking paradigm may miss important aspects of cognitive development that could only be revealed under more "active" eye-tracking conditions.

Corbetta, Guan, and Williams (in press) summarize two ways in which the study of what Rao and Ballard (1995) called "active vision" can be accomplished with infants. The first is the most obvious—replace the spatially fixed eye-tracker with a head-mounted eye-tracker. This has been a method used with adults for nearly 20 years, but until recently the size (and weight) of such systems precluded their use with infants. Corbetta's laboratory was one of the first to successfully collect eye-movement data from infants wearing a head-mounted tracker, and even smaller trackers have now been deployed in a few laboratories (Franchak, Kretch, Soska, & Adolph, 2010; Kidd & Aslin, 2011; Yu, Smith, Fricker, Xu, & Favata, 2011). As noted by Corbetta et al., a head-mounted system consists of a

scene camera (pointed outward from the middle of the infant's forehead) and a pupil camera (pointed toward one of the infant's eyes). In the system used in Corbetta's laboratory, the pupil camera reflects off a half-silvered mirror positioned in front of one eye. In other systems, the pupil camera is held in place by a curved stalk to enable a direct view of the eye. Both of these arrangements can be distracting to the infant, especially when first positioned on their head, and of course the mirror or stalk serves as a target for infant reaching. These systems require some subtle adjustments, in situ, to align the pupil camera, drawing the infant's attention (sometimes repeatedly) to the mirror or stalk. But with suitable distractions (e.g., colorful toys or soap bubbles), most infants can be fitted and successfully calibrated using procedures similar to remote (fixed in place) eye-trackers.

Head-mounted eye-trackers enable researchers to ask a variety of questions that could not be addressed with a remote eye-tracker. Corbetta's laboratory is interested in visual control of motor behaviors (see also Franchak et al., 2010), and the infant's gaze can be captured as they make head movements to look anywhere in their immediate surroundings. There are some limitations to these head-mounted systems. First, because the scene viewed by the infant, and documented in the scene camera's image, is under the control of the infant, it is not possible to use automated coding algorithms as with remote (spatially fixed) eye-trackers. Thus, all of the gaze data must be hand-coded by human observers who view the dynamic scene camera video with gaze position (typically indicated by cross-hairs) superimposed on a frame-by-frame basis. Second, the view of the infant's surroundings provided by the scene camera has a single angular dimension with respect to the forehead. If this angle is adjusted for distant objects, then any looking to the infant's hands (in near space) is outside of this field of view. If the angle is adjusted for near space, then distant objects are not in view. Perhaps in the future, head-mounted systems will employ two scene cameras, each with separate calibrations, to enable a larger vertical dimension of the infant's surroundings to be captured.

Corbetta et al. (in press) created a clever way of overcoming this limitation of having to choose between a near-view or a far-view scene camera. Objects were placed in an opening that replaced the typical LCD screen of a spatially fixed (remote) eye-tracker. The infant was positioned so that the objects were out of reach but their eye movements could be recorded. Then, once the infant had the opportunity to look at the objects, the infant was moved closer to the objects so that they could choose to reach for them. In this way, the object of regard could be determined prior to the infant's reach. Although this paradigm does not address whether infants rely on visual information about the object *during* the reach (because the eye-track could not be maintained when the infant was moved closer to the objects), it

did reveal a very interesting finding about how infants direct their attention to the target of their eventual reaching response. The earliest reaching responses at approximately 4 months of age were only rarely preceded by looking at the specific location on the object where the reach would subsequently occur (e.g., the handle on a cup). By 9 months of age, however, infants almost always directed their pre-reaching gaze to the specific location where the reach made contact with the object. These results suggest that vision is more reliably a key component in the control of reaching as infants gain experience with objects, a finding that could only have been discovered by using an eye-tracker.

Unfortunately, eye-tracking data do not always provide clear insights about infant cognition, either because we have not determined the relevant dependent measure or because the data do not, in principle, reflect an underlying cognitive process in the specific testing situation used. For example, based on data from adults, we make the assumption that where infants look is directly related to what they are processing about a stimulus. But we know that this assumption is only correct to a first approximation because adults, and presumably infants, also process information that enters the visual system via the peripheral retina. For example, we negotiate doorways and steps without looking in detail at the doorframe or the floor, and we are sensitive to the entrance and exit of people in our field of view despite never looking at them directly.

Sirois and Jackson (in press) document in a variant of Baillargeon's "drawbridge" paradigm that looking times do not match classic findings of object permanence in 10-month-olds. Presumably, the violation-of-expectancy paradigm used by Sirois and Jackson, identical to the 2-D paradigm used by Cashon and Cohen (2000) and similar to the 3-D paradigm used by Baillargeon, Spelke, and Wasserman (1985), does not provide a global looking-time metric of object permanence. Moreover, Sirois and Jackson could find neither a macrostructure nor a microstructure in their eye-tracker data to account for this discrepancy between Baillargeon et al. and Cashon and Cohen. Given that 10-month-olds show object permanence using a classic manual search task, the conclusion is not that object permanence is absent, but rather than the specific paradigm used and/or the dependent measure is not a robust indicator of object permanence.

What other measure could one obtain to overcome this apparent insensitivity of eye-tracking data? Sirois and Jackson (in press) provide evidence that pupil size is a more sensitive measure of underlying cognitive processes. Because pupil size is recorded as part of any video-based eye-tracking system, it is a "free" dependent variable. Unfortunately, not everything that is free is worth the price. Although Sirois and Jackson provide evidence that the time-course of mean pupil size (across trials and infants) varies by

stimulus condition (possible versus impossible object-occlusion events), they did not regress out the effect of stimulus luminance, which varies by condition (the bright drawbridge is present longer in the possible occlusion event). Given the fact that they have both gaze position and pupil size data on a trial-by-trial basis, they could use as a predictor in a mixed-model regression whether the infant was fixating the drawbridge (thereby causing a pupil constriction). If after compensating for this luminance effect the residual measure of pupil size were still a significant contributor to the condition effect, then their results would have been more convincing.

Gredebäck, Eriksson, Schmitow, Laeng, and Stenberg (in press) provide data from 14-month-olds' fixations of facial stimuli varying in emotional expressions that eye-tracking may be less sensitive than pupil size. At issue is the role of early experience with faces, specifically whether the primary caregiver was a woman or a man. Unfortunately, there is a societal asymmetry here, and the two groups studied by Gredebäck et al. were either exclusively raised by the mother or more balanced between the father and mother (but still mother-biased). In studies of face scanning using eye-trackers (e.g., Hunnius & Geuze, 2004), there are a variety of dependent measures. Gredebäck et al. chose to focus on two: the number of fixations and the root mean square (RMS) variance in gaze. Curiously, they did not report the *duration* of fixation to the faces depicting different emotional expressions (as a proportion of total on-screen looking). RMS is a measure of the dispersion of overall scanning and is similar to the less precisely defined *broad* versus *narrow* scanpaths of faces described by Maurer and Salapatek (1976). This RMS measure could have implications for face processing, but Gredebäck et al. only assessed how infants scanned the faces, not whether that scanning affected their discrimination or memory for the faces. Similarly, the measure of *number* of fixations has no obvious linking hypothesis to face processing—do short looks imply more engaged processing (as suggested by Bronson, 1991), and by implication suggesting that more fixations are associated with better processing? Without a measure of processing we cannot know the answer to these questions.

Given the generally negative results obtained by Gredebäck et al. (in press) using eye-tracking (except for larger RMS values by two-parent reared infants), they ask whether pupil size is a better predictor of infants' processing of facial expressions (and whether responding to the gender of the face stimuli is influenced by experience with female and male caregivers). Although their results suggest a complicated set of interactions between gender, rearing condition, and type of face expression, the main finding of larger pupil sizes to neutral facial expressions depicted by someone other than the primary caregiver is not grounded in a specific linking hypothesis. What is the relationship between tonic pupil size (not the more traditional *change* in

pupil size) to emotional expressions depicted in photographs? Does this imply preference, arousal, or negative affect? Because pupil size is jointly determined by low-level stimulus factors (e.g., luminance) as well as by a balance between the parasympathetic and sympathetic nervous system pathways, it seems prudent to attempt to gain a better handle on how these underlying mechanisms interact before drawing strong inferences about cognitive processes. What is needed to validate the use of tonic pupil size in studies of infant cognition, therefore, is to take a continuum of stimuli that vary along a single dimension (e.g., facelike versus non-facelike) that is unconfounded by low-level factors. The goal is to determine whether tonic pupil size maps monotonically onto looking-time preferences for stimuli along this dimension, preferably within infants and not just across groups of infants.

CONCLUDING REMARKS

In this commentary, I have attempted to do three things. First, it is important to pay close attention to the technical details of how an eye-tracker works and the many ways in which it can fail to provide the kind of reliable and valid data we all expect in recording gaze from infants. No eye-tracker should be used "out of the box" without verifying that it meets the standards for spatial and temporal accuracy that are required to draw conclusions from your data. Although eye-trackers have their limitations, when used appropriately—recognizing that the quality of the data must match (or exceed) the requirements for addressing a particular research question—they have the potential to provide new insights about cognitive development.

Second, eye-trackers are not only seductively easy to use but they generate voluminous amounts of data. It is perfectly fine to conduct exploratory research in an attempt to find potentially fruitful dependent measures. But measures that are correlated with variations in a stimulus do not explain the cognitive mechanism underlying how those stimuli are processed. One must follow exploratory research with experiments that manipulate a crucial stimulus variable and test a linking hypothesis between the dependent and independent variables. There are many exciting data-mining techniques being applied to eye-tracking data, and the iterative process of exploring new techniques and generating new hypotheses will likely lead to more sophisticated models of cognition in infants.

Third, depending on the research question, spatially fixed (remote) versus head-mounted eye-trackers may provide the better platform for addressing a particular research question. By clever experimental design and technical advances in miniaturization, we can expect the use of both types of

eye-trackers to continue to become a common method in studies of infant cognition. A final step in this process may, one day, be the use of virtual reality displays with infants, which would enable the experimenter to control the visual world even as the infant makes active movements.

ACKNOWLEDGMENT

Preparation of this article was supported, in part, by NIH grant HD-37082.

REFERENCES

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*, 48–53.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*, 191–208.

Bronson, G. W. (1991). Individual differences in rate of visual encoding. *Child Development*, *62*, 44–54.

Cashon, C. H., & Cohen, L. B. (2000). Eight-month-old infants' perceptions of possible and impossible events. *Infancy*, *1*, 429–446.

Corbetta, D., Guan, G., & Williams, J. L. (in press). Infant eye-tracking in the context of goal-directed actions. *Infancy*, DOI: 10.1111/j.1532-7078.2011.00093.x.

Dodge, R. (1907). An experimental study of visual fixation. *Psychological Review*, *35*, 1–95.

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, *72*, 1003–1015.

Franchak, J. M., Kretch, K.S., Soska, K.C., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research*, *50*, 2766–2774.

Gredebäck, G., Eriksson, M., Schmitow, C., Laeng, B., & Stenberg, G. (in press). Individual differences in face processing: Infants' scanning patterns and pupil dilations are influenced by the distribution of parental leave. *Infancy*, DOI: 10.1111/j.1532-7078.2011.00091.x.

von Hofsten, C., & Rosander, K. (1997). Development of smooth pursuit tracking in young infants. *Vision Research*, *37*, 1799–1810.

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*, *6*, 231–255.

Kidd, C., & Aslin, R. N. (2011, March). *Learning at a glance: Infants' attention to rapid parental social cues for lexical references*. Paper presented at the biennial meeting of the Society for Research in Child Development, Montreal, Canada.

Maurer, D., & Salapatek, P. (1976). Developmental changes in the scanning of faces by young infants. *Child Development*, *47*, 523–527.

McGinnis, J. M. (1930). Eye movements and optic nystagmus in early infancy. *Genetic Psychology Monographs*, *8*, 321–430.

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, *6*, 203–229.

Morgante, J. D., Zolfaghari, R., & Johnson, S. P. (in press). A critical test of temporal and spatial accuracy of the Tobii T60XL eye tracker. *Infancy*, DOI: 10.1111/j.1532-7078.2011.00089.x.

Oakes, L. (2010). Infancy guidelines for publishing eye-tracking data. *Infancy*, *15*, 1–5.
Rao, R. P. N., & Ballard, D. H. (1995). An active vision architecture based on iconic representations. *Artificial Intelligence*, *78*, 461–505.
Salapatek, P., & Kessen, W. (1966). Visual scanning of triangles by the human newborn. *Journal of Experimental Child Psychology*, *3*, 155–167.
Shukla, M., Wen, J., White, K. S., & Aslin, R. N. (2011). SMART-T: A system for fully automated anticipatory eye-tracking paradigms. *Behavior Research Methods*, *43*, 384–398.
Sirois, S., & Jackson, I. R. (in press). Pupil dilation and object permanence in infants. *Infancy*, DOI: 10.1111/j.1532-7078.2011.00096.x.
Yarbus, A. L. (1967). *Eye Movements and Vision*. New York: Plenum.
Yu, C., Smith, L. B., Fricker, D., Xu, L., & Favata, A. (2011, March). *What joint attention is made of—a dual eye tracking study of child-parent interaction*. Paper presented at the biennial meeting of the Society for Research in Child Development, Montreal, Canada.
Yu, C., Yurovsky, D., & Xu, T. (in press). Visual data mining: An exploratory approach to analyzing temporal patterns of eye movements. *Infancy*, DOI: 10.1111/j.1532-7078.2011.00095.x.

APPENDIX

How to assess timing accuracy?

The flow of information in a typical eye-tracking experiment with infants is quite simple: A visual stimulus appears on the display screen and x, y eye-position coordinates are sampled and stored on disk. How can something so simple go wrong? The problem comes from the passage of information between various components of the display and eye-tracker. These signals can only be passed reliably from one component to the next if the hardware and software are keeping track of timing events with the same "clock."

Consider the case of a "go" event to start a trial, which consists of a ball appearing on the display screen and moving back and forth behind an occluder. When this event is triggered (either via a key-press or inside a software package such as E-Prime or Tobii Studio), there can be delays in (1) when the display actually appears on the screen and (2) when the video file begins to play. If these delays are not taken into account, then the x, y coordinates from the eye-tracker are not synchronized with what the infant is actually seeing. Importantly, any delay will appear to be an *anticipation* by the infant, even if they are actually responding *after* the stimulus event.

But there is a second source of timing errors that resides within the eye-tracker itself. When the eye moves, the eye-tracker hardware and software must compute an estimate (based on current data and prior calibration data) of the x, y coordinates. This computation has a delay, which is constant (e.g., 10 msec) but can also be variable (e.g., 10 msec at one time and 100 msec at another time). Both delays must be known to accurately reflect the relationship between where the infant is looking at that moment in time and where the eye-tracker says the infant is looking at that moment in time.

What is needed to assess these different types of delays is an external device that records when the stimulus appeared (or began to move on the display) and when the eye actually moved, and then to relate these timing parameters to what the eye-tracker claims about these events. This can be accomplished with three pieces of equipment: a high-speed camera, a digital video recorder, and a mirror. The key is to use the high-speed camera and the digital video recorder to simultaneously capture an image of the stimulus display and an image of the eyes. This can be done by placing the high-speed camera behind the shoulder of an adult who is aligned with the eye-tracker display and who has already been calibrated. The high-speed camera captures an image of the display screen and therefore can see when a stimulus moves from one position on the display to another position. By placing a small mirror to the side of the display, but visible by the high-speed camera, the digital video recorded from the high-speed camera contains *in the same image* both the movement of the stimulus and the movement of the adult's eyes reflected in the mirror.

From the digital video, therefore, one can examine on a frame-by-frame basis when the stimulus moved on the display screen and when the eyes of the adult moved in response to this stimulus movement. The high-speed camera should have a sampling rate that is well above the rate used by the eye-tracker. For example, if the eye-tracker has a sample rate of 60 frames per sec, the high-speed camera should have a sample rate of 300 frames per sec, thereby allowing the coding of timing errors that are a fraction of the eye-tracker's sample rate. By implementing a series of stimulus movements and eye movements by the adult, one can easily compute the latency of the eye movement relative to the stimulus movement, as well as the time *between* each pair of movements. Importantly, these timing measures are unaffected by any delays introduced by the computer creating the stimulus display or the eye-tracker hardware or software.

Given these objective measurements of how the adult is actually responding to the stimulus, one can then ask whether the data provided by the eye-tracker correspond to these timing parameters. On a trial-by-trial basis, one can determine from the digital video recording when the stimulus moved on the screen and how long it took the eye movement to respond to this stimulus movement. Then from the output of the eye-tracker, one can examine these same timing events to determine the errors introduced by the eye-tracker hardware and software, as well as by any other software (e.g., E-Prime) and hardware (e.g., video cards and Ethernet connections) used to communicate between the eye-tracker and its host (and satellite) computers. From these comparisons, one can obtain the average timing delay and the trial-by-trial variability in these timing delays.