A behavioral method for efficient screening of visual acuity in young infants

I. Preliminary laboratory development

Velma Dobson, Davida Y. Teller, Clifton Ping Lee, and Brenda Wade

A technique for rapid behavioral screening of grating acuity in infants 1 to 4 months of age is described. The approach, called forced-choice preferential looking (FPL), depends upon the fact that normal infants will stare fixedly at acuity gratings with stripe widths above a rather abrupt cutoff width. The present task is to find the minimum stripe width, termed the diagnostic stripe width, to which infants with normal visual acuity will readily respond. The expectation is that infants with below-normal acuity will not be able to respond to diagnostic stripes so defined. To assess the feasibility of such a test procedure and define preliminary diagnostic stripe widths for infants 4, 8, 12, and 16 weeks of age, 76 presumptively normal infants were tested in the laboratory with a 40-trial FPL procedure. Sixty-nine infants (91%) completed the test procedure. For each age group, a preliminary estimate of the diagnostic stripe width was found. The group performance was uniformly high for the stripe widths designated as diagnostic and fell off sharply for finer stripes, confirming the feasibility of the approach.

Key words: visual acuity, screening, human infants

Methods for assessment of visual acuity in infants would be useful to pediatricians and ophthalmologists for early detection of visual problems. Until recently, however, neither normative infant acuity data nor clinically applicable techniques for measuring acuity in infants have been available.

Within the past few years, a considerable number of studies of the course of development of visual acuity in infants between birth and 6 months of age have appeared. These studies have used optokinetic nystagmus (OKN), preferential looking (PL), and the visually evoked potential (VEP) to measure acuity. Data accumulating from many laboratories suggest that visual acuity increases regularly with age over at least the first 4 to 6 months postnatal. Data from different laboratories employing variants of the same technique show good agreement on age norms and age trends for group data. In addition, there is relatively low variability in acuity values obtained from extensively tested individual infants of a given age. Thus, reliable normative descriptions of the development of visual acuity are rapidly becoming available. (For reviews see Dobson and Teller.)

Recent publications have indicated that each of the three techniques which have been used to gather normative data in the laboratory can be used for clinical assessment of acuity in infants. Enoch and Rabinowicz.
used OKN to follow the development of visual acuity in an infant subsequent to removal of a monocular cataract. They found that acuity development was arrested in the aphakic eye but began to develop along a more or less normal time course after the fitting of a correcting contact lens. Sokol9 has used the VEP to assess visual acuity in two infants seen clinically. In one infant, who had a monocular cataract, the VEP was found to be reduced in the affected eye. A second infant, born 8 weeks prior to term and tested at 22 weeks postnatal, exhibited VEP responses more similar to those of a full-term 18-week-old infant than a full-term 22-week infant. Miranda19 and Fantz et al. 20 used the PL technique to show that the development of acuity in a group of premature infants lagged behind the development of acuity in a group of term infants, with a time course predictable on the basis of conceptional age. In addition, performance on a PL task has been shown to be among the best available predictors of later neurologic abnormalities in high-risk neonates.21 Thus, there is ample reason to attempt development of techniques applicable to the screening of visually at-risk infants.

The present article presents the conceptual framework and preliminary laboratory results of a modification of the PL procedure intended for clinical use. The article that follows presents the results of the first clinical trials using the procedure.

Materials and methods

General procedure. A modification of the original preferential looking (PL) technique has been developed in our laboratory and used to obtain psychophysical estimates of acuity in individual infants. In this forced-choice preferential looking (FPL) procedure, an infant is held in front of a display screen (Fig. 1). A striped acuity grating is presented in one of two possible positions, either at the left or at the right of the screen. An observer, uninformed as to the position of the test grating, is located behind a peephole at the center of the screen. The observer’s task is to judge the position of the grating pattern on each trial on the basis of the infant’s looking behavior (Fig. 2). Each infant is tested extensively, with 20 or more trials on each of four or five different stripe widths. The percent of trials in which the observer correctly judges the location of the striped pattern can then be plotted as a function of the stripe width, to yield a psychometric function (Fig. 3). Acuity is estimated as some point on the psychometric function, e.g., 75% correct by the observer. Thus, the FPL procedure results in an estimate of acuity for each infant.

However, two features of this procedure render it as yet impractical for clinical use. First, even if the stripe widths needed to span the function can be guessed ahead of time, as they now can be for normal infants, a substantial number of trials at...
Fig. 2. Observer’s view of the infant through the peephole. A, Acuity grating to the observer’s left. B, Acuity grating to the observer’s right.

Fig. 3. Psychometric function for an 8-week-old infant tested with the FPL procedure. The observer’s percent correct is plotted as a function of the five stripe widths on which the infant was tested. Conversions to Snellen equivalents (using the convention that 1' per stripe equals 20/20 Snellen) are given on the abscissa. The infant performed near 100% for 40' and 20' gratings but near chance for 10' and 5' targets, with an estimated “threshold” (75%) at about 16' or 20/320 acuity.

several different stripe widths is required to obtain a well-defined psychometric function. To obtain a psychometric function with 20 trials per point at five stripe widths can take from 2 to 4 hours of testing, depending upon the infant. Second, if the stripe widths needed to span the function cannot be guessed a priori—and, almost by definition, they cannot be for clinically interesting infants—then even more stripe widths must be used. Even with a highly efficient search strategy, many stripe widths would have to be sampled to find and define the function if it were in an unpredictable location.

A procedure which should on average be less time consuming would be to test an infant at a single point on the stripe-width continuum and then use the infant’s score on that stripe width as a diagnostic indicator of whether the infant’s acuity.
is normal or not. Ideally, such a diagnostic stripe width would be selected to be large enough so that virtually all infants with normal visual acuity could easily detect the stripes but small enough that virtually all infants with abnormally low acuity would fail to detect them. For clinical evaluations of individual infants, in addition to using a minimal number of trials, it is also necessary to achieve statistically reliable results for each infant.

With these strategies in mind, we have defined the diagnostic stripe width to be the narrowest stripe width on which 95% of normal infants of a given age will get 15 or more out of 20 trials correct; i.e., 75% or more correct on 20 trials.* This value was chosen because with \( p = 0.5 \) and \( n = 20 \), 75% correct is statistically significantly above chance at the 0.05 level (exact binomial test).

The present task, then, was to find and validate diagnostic stripe widths for normal infants of various ages.

**Apparatus.** The apparatus was identical to that described by Teller et al. and is shown in Fig. 1. Basically, the apparatus consisted of a gray screen (Crescent Cardboard No. 651) in which two 9 cm stimulus holes and a 4 mm peephole were cut. On each trial, an acuity grating (Biometrics Division, Narco-Biosystems) was placed behind one stimulus hole and a piece of the gray cardboard was placed behind the other. The acuity gratings and gray cards were embedded in a wheel mounted behind the cardboard screen so that the position of the stripes (left vs. right) and the size of the stripes (control stripes vs. test stripes) could be changed rapidly during the test session.

The acuity gratings had a nominal 1:1 duty cycle and a nominal contrast of 52%. The space-average luminance of the individual striped cards ranged from 1.25 to 1.30 log cd/m\(^2\) (measured with an International Light IL500 Photometer) and approximately equaled the luminance of the gray cardboard stimulus and screen (1.23 log cd/m\(^2\)).

*It must be emphasized that the diagnostic stripe width is not itself an absolute measure of acuity, but falls about 1.5 octaves below the FPL acuity at each age. (See Dobson and Teller.)

Stripes widths of the grating stimulus were 0.95, 0.48, 0.32, 0.24, and 0.12 cm, which correspond to approximately 80', 40', 27', 20', and 10' of visual angle, respectively, for an infant held at the test distance of 36 ± 3 cm from the centers of the two stimulus positions. Conversions to Snellen equivalents are given in Fig. 3.

Three adults are required to test an infant: the **observer**, who observes the infant and judges the position of the striped card on each trial; the **holder**, who holds the infant but cannot see the stimulus display (Fig. 1); and the **experimenter**, who sets up the trials, records the observer’s judgments, and provides trial-by-trial feedback to the observer.

**Subjects.** Subjects were solicited by letters to parents listed in the birth announcements section of the newspaper and were paid $5.00 for the test session. Fourteen 4-week-old, twenty-one 8-week-old, nineteen 12-week-old, and twenty-two 16-week-old infants were tested. All infants were born within 10 days of their due date, as reported by the parents, and each was tested at a single age, within 3 days of his or her fourth-, eighth-, twelfth-, or sixteenth-week birthday.

**Selection of stimuli.** The goal of the study was to estimate, to one octave* or better, the diagnostic stripe width, i.e., the smallest stripe width on which 95% of infants with normal visual acuity would perform significantly above chance. Therefore, for each age it was necessary (1) to find a stripe width on which virtually all infants would perform significantly above chance in 20 trials and (2) to show that a substantial fraction of normal infants would fail to perform significantly above chance on a stripe width finer by 1 octave or less. Intuition and previous experience in testing infants led us to use 40' and 20', 27' and 20', 20' and 10', and 20' and 10' test stripes for 4-, 8-, 12-, and 16-week infants, respectively. Each infant was tested with only one of the two test stripe widths selected for his or her age group.

In addition to the test stripes, every infant was tested with a set of large, readily resolvable control stripes, of 80' subtense. The control stripes were intended to serve two purposes. First, they helped keep the infant interested in the procedure. Second, they indicated whether poor performance by an infant on the test stripes (if it occurred) was due to a behavioral problem, e.g., fussiness or drowsiness, or to limited acuity. High

*An octave is a halving or doubling of stripe width or spatial frequency, and a halving or doubling of the denominator in Snellen notation (Fig. 3).
Fig. 4. Scoresheet containing one of the orders of trials used for testing an infant with the diagnostic stripe width. On the first trials, large control stripes are presented to allow the infant to become familiar with the procedure and the observer to become familiar with the infant's looking pattern. The final 30 trials are a mixture of 10 control stripes and 20 diagnostic stripes. The size of the diagnostic stripe width used depends on the age of the infant. In the present study, all 40 trials were used. Arrows indicate a set of possible cutoff points of termination of testing for increased efficiency. (See discussion of rescoring.)

performance (≥75% correct) on the part of the observer on both test and control stripes was taken to indicate both that the infant was testable and that the test stripes could be resolved. Poor performance on the test stripes but high performance on the control stripes was taken to indicate a testable infant unable to resolve the test stripes. Poor performance on both test and control stripes (should it occur) would be taken to indicate either an untestable infant or one with severely limited acuity.

Procedure. Each infant received 40 trials using the FPL procedure. The time required to complete the 40 trials was recorded. A sample score
Infant visual acuity screening. I. Laboratory

FIG. 5. Percent correct on test and control stripes, for infants tested in preliminary estimation of diagnostic stripe widths. Scores of each infant on a test stripe width and on the 80' control stripes are shown. Preliminary diagnostic stripe widths are taken to be 40', 27', 20', and 20' for 4-, 8-, 12-, and 16-week infants, respectively. The three symbols representing scores on the control stripes indicate scores of infants who passed (scored ≥ 75%) on either of the two test stripes (•) and scores of infants who failed (scored < 75% correct) on the smaller (○) or larger (×) of the test stripes at a particular age.

Sheet is shown in Fig. 4. For the first 10 trials, the control stripes were used. These served to acquaint the infant with the procedure and allowed the observer to become familiar with the infant’s looking pattern. Following these trials, the infant received 30 trials in which 10 of the control stripes were intermixed quasirandomly with 20 trials of the test stripes being used for that infant. On each trial, the observer’s task was to say whether the stripes appeared on the left or on the right side of the screen. Trial length was variable, depending upon the length of time the observer felt was required to make a judgment. Typically, trial length was between 10 and 30 sec, although occasional trials required less than 5 or more than 60 sec.

An infant’s results were designated pass if he or she scored 75% or better on the 20 test trials, and fail if the score fell below 75%.

Testing on each potential diagnostic stripe width was continued either until 10 infants had been tested with that stripe width (of whom at least eight had passed) or until three infants had failed. In the latter case, no more infants were tested at that stripe width, since it was unlikely to be the diagnostic stripe width.

Results

The scores obtained by each infant on test and control stripes are shown in Fig. 5. The abruptness of the change in the infants’ behavior with changes in stripe width suggests that the concept of a diagnostic stripe width is a meaningful one. For example, for the 4-week-old infants (Fig. 5), three out of three infants tested with 20' stripes failed (scored less than 75% correct), while 10 out of 10 infants tested with 40' stripes passed (scored greater than 75% correct). These data, although limited by a small sample size, suggest that the diagnostic stripe width for 4-week-old infants falls between 20' and 40'.
suggest diagnostic stripes between 20' and 27' for 8-week infants and stripes between 10' and 20' for 12- and 16-week infants. Therefore, for our preliminary clinical trials, we used 40' stripes for infants less than 7 weeks of age, 27' stripes for 8- through 11-week infants, and 20' stripes for infants 12 through 16 weeks of age.22

The results on the control stripes for the 4-, 8-, and 12-week infants also support the efficacy of the test procedure. For these age groups, virtually all infants scored 85% or above on the control stripes (mean = 95%). This performance suggests that any failure to perform above chance on the test stripes was probably due to real acuity limitations rather than insensitivity of the procedure. One 12-week infant performed below 75% on the control stripes; however, the fact that this infant scored 95% on the 20' stripes suggests that the test was an adequate assessment of that infant's visual acuity.

Sixteen-week-old infants are, on the average, a little harder to test. Some of these infants became fussy toward the end of testing and objected to being held in front of the screen. The difficulties encountered in testing this age group are reflected in the relatively low scores of the 16-week infants on the control stripes (mean = 89%). Nevertheless, two of the three 16-week infants who failed with the 10' stripes performed well on the 80' stripes, suggesting that the test adequately assessed their detection of the test stripes. The third performed poorly on both, suggesting lack of testability. The one 16-week infant who failed on the 20' stripes scored only 80% on the 80' stripes. By strict usage of our scoring rules, he would be diagnosed as having failed on the basis of poor acuity. However, his low performance coincided with fussiness at the end of testing. Diagnosis of limited acuity vs. untestability would in fact have been unclear for this infant, and retesting on a later date would be indicated. (See also the results of rescoring, below.)

Seven of seventy-five (9%) of the infants failed to complete the test session. Sleepiness prevented completion of testing of five infants, evenly distributed across ages. Two 16-week infants did not complete testing due to fussiness. For the 69 infants who completed all 40 test trials, the average time required was 24 mins (SD = 10 min, range = 10 to 45 min).

Recoring. It was often obvious to the observer that infants who were going to pass could see the diagnostic stripes long before the end of the 40-trial procedure. Thus, it seemed sensible to stop testing each infant as soon as he or she obtained a score statistically significantly above chance. Binomial probabilities indicate that the probability that an infant who cannot see the diagnostic stripes could obtain a score of 5/5, 7/8, 9/11, 11/14, 12/16, or 14/19 is less than 0.04, with a cumulative probability of 0.088.22 If it is desired to keep the cumulative probability at 0.05, the cutoff criteria 5/5, 9/10, 13/15, and 15/20 may be used, for a cumulative probability of 0.050.

Rescoring the data using either set of cutoff criteria showed that of the 46 infants who scored significantly above chance on the 40-trial procedure, all would have passed, and the average number of trials required to complete testing in these infants would have been 20 instead of 40, with a concomitant reduction of testing time. Of the 13 infants who failed to score significantly above chance on the 40-trial procedure, only one (the 16-week-old infant who failed the 20 min stripes) would have obtained a passing score if either set of cutoff criteria had been used. Since this infant's poor score on the 40-trial procedure was primarily due to fussiness during the latter part of the test, the results using the cutoff criteria are probably a more accurate indication of the infant's abilities than the results after 40 trials. Furthermore, five of the seven infants who were eliminated from the 40-trial procedure due to sleepiness or fussiness during testing would have been able to complete the version of the procedure containing the cutoff criteria. Thus, use of cutoff criteria to shorten the 40-trial procedure does not appear to change the conclusions drawn about an infant's vision in the laboratory; it merely shortens test time to one half or less.
for infants with good visual acuity and allows testing of infants who would not tolerate a longer 40-trial procedure.

Discussion

The results of the present study suggest that it is possible to find a reliable minimum (diagnostic) stripe width on which most infants of a particular age with presumptively normal visual acuity can quickly perform significantly above chance during testing with the FPL procedure. For each age group, infants tested with the stripe width later designated as diagnostic obtained uniformly high scores, while infants tested with stripes 0.5 or 1 octave narrower showed a much greater variability of scores.

On the basis of previous laboratory acuity data on infants, it was possible to predict fairly accurately which stripes would be the diagnostic stripes. This success suggests that our quick assessment taps a visual function very similar to that measured in more extensive FPL acuity testing. The validity of our results is further supported by the parallel fashion in which the size of the diagnostic stripe widths and previously reported estimates of PL acuity change with increasing age of the infant. Further, the test is relatively quickly accomplished, and more than 90% of the infants tested completed testing. Thus, the technique exhibits some of the properties of efficiency and validity necessary for eventual usefulness in the clinical setting.

Further development of the procedure will include several aspects. First, the size and diversity of the sample on which the diagnostic stripes are selected will be increased, and finer estimates of the diagnostic stripe widths will be made. Second, increasingly shorter and more efficient forms of the procedure are being developed, with the goal of reducing the time required for screening acuity to a maximum of 10 min. Third, monocular testing of infants would presumably be of considerable value in early detection of monocular problems. We are presently attempting to test infants monocularly. Fourth, infants are being tested in the Department of Ophthalmology at Children's Hospital Medical Center, Boston, in order to try the technique in a clinical as well as a laboratory setting. We wish to see whether infants at risk for poor visual acuity fail to perform significantly above chance on the diagnostic stripes, and whether infants who perform poorly without prior diagnosis of risk will be found to have significant visual problems.

Finally, it should be reemphasized that the present modification of the FPL technique is specialized as a screening test and not as a measure of an infant's acuity per se. That is, each infant is tested against a single age norm—the diagnostic stripes appropriate for his or her age—and simply passes or fails; no estimate of acuity is made. Acuity estimates can be made with FPL by using several stripe widths in random order on the same infant and generating a psychometric function similar to that in Fig. 3, as in our usual laboratory procedure, or by the use of staircase or staircaselike procedures. Attempts are in progress in our laboratory and in the laboratories of others to develop a preferential looking technique that is efficient and reliable enough to yield useful acuity estimates for individual infants in clinical settings.

We thank Michelle Backholm and Michael Sekel for assistance in running subjects, and Marjorie Zachow for secretarial assistance.

REFERENCES

7. Marg, E., Freeman, D. N., Peltzman, P., and...


