Vision Res. Vol. 22, pp. 1025 to 1031, 1982
Printed in Great Britain.

0042-6989/92 $3.00 + 0.00
Pergamon Press Ltd

A FAILURE TO OBSERVE NEGATIVE PREFERENCE IN INFANT ACUITY TESTING

MARTIN S. BANKS, BENJAMIN R. STEPHENS and JAMES L. DANNEMILLER
Department of Psychology, Mezes Hall 330, The University of Texas at Austin, Austin, TX 78712, U.S.A.

(*Received* 10 *April* 1981; in *revised form* 7 *December* 1981)

Abstract—Held *et al.* (1979) Vision *Res. 19,* 1377-1379, reported that infants' psychometric functions in preferential looking experiments exhibit a region of below chance performance or "negative preference". They argued from this that previous preferential looking experiments may have systematically underestimated acuity because they ignored this negative preference. We present data from an experiment designed to reveal negative preference in the version of the preferential looking paradigm used in our laboratory. The results do not exhibit negative preference apart from random variations about 50% Gwiazda *et al.* (1980) *Am. J. Optom. physiol. Opt. 57,* 420-427; 428-432, developed a psychophysical procedure for infant testing that is designed to yield accurate threshold estimates in a short period of time. We present the results of computer simulations of more conventional psychophysical procedures. The simulations demonstrate that an up-down staircase procedure is more efficient than the Gwiazda *et al.* procedure.

INTRODUCTION

The preferential looking technique has become widely used to test infant visual acuity. Held *et al.* (1979) recently reported data which call one of the technique's basic assumptions into question. This paper and the adjoining one (Teller *et al.* 1982) concern the criticisms raised in the Held *et al.* report.

Held *et al.* reported that infants' psychometric functions in preferential looking experiments are nonmonotonic-, that is, with increasing spatial frequency such functions drop from high levels of preference for the grating (approximately 100010) to *below* 50% before finally asymptoting at 50% preference. We will call this region of below chance performance the negative preference dip. Held *et al.* argued correctly that performance significantly below chance implies discrimination of the grating and blank fields and hence detection of the grating. Because they have ignored preferences significantly below 50%, they concluded that they have "systematically underestimated acuity". Moreover, they "suspect this conclusion is also true of other preference procedures..." (p. 1378).

Held's group has developed a psychophysical procedure intended to yield accurate threshold estimates in a short period of time (Gwiazda *et al.*, 1980b). They claimed that the procedure is quite efficient in studies of normal and clinical populations. The procedure's reliability is increased significantly by the presence of a negative dip in the infant's psychometric function (Gwiazda *et al.*, 1980b).

This paper makes two specific points concerning the results and procedures described by Held *et al.* (1979) and Gwiazda *et al.* (1980b). First, we present data from an experiment specifically designed to reveal negative preference dips in a conventional preferential looking experiment. We used the methodology currently employed in our laboratory. It is similar but not identical to that of Held *et al.* Our motivation was simply to determine if negative preference occurs with the methodology, apparatus, and age ranges we typically use in our laboratory. The results do not exhibit a negative preference dip apart from random variation about 50%. Second, we present the results of computer simulations of conventional up-down staircase procedures and of the Gwiazda *et* al. procedure. The simulations indicate that up-down procedures are more efficient; that is, for a given number of trials, the threshold estimates of up-down procedures are less variable and biased than those of the Gwiazda *et al.* procedure. The simulations also indicate that any procedure using small numbers of trials yields relatively unreliable estimates of threshold. Consequently, only substantial differences in acuity can be detected with confidence.

METHODS

Infants were recruited by letter and telephone. We were primarily concerned with young infants, so only 1- to 3-month olds were tested (mean age = 71.0 days, SD = 19.3, range = 38-108). Eleven of the 17 infants who participated provided complete data sets. The others did not complete the experiment due to fussiness, sleepiness, or scheduling difficulty.

The forced-choice preferential looking technique was employed (Teller, 1979). Stimuli were generated on a large-screen CRT (Hewlett-Packard 1317A) using the method of Campbell and Green (1965). At the 40 cm viewing distance, the display subtended 48º x 37º. Space-average luminance was 10.6cd/m^2 and the surround was dark. Simultaneous presentation of a squarewave grating and a uniform held was accomplished by splitting the display screen electronically at midline. Thus, the grating and blank field were adjacent and equal in space-average luminance, spectral composition, and size. The contrast of the squarewave grating was 0.80[*]. The range of spatial frequencies used was carefully chosen to detect the presence of any negative preference dip. Based on infant contrast sensitivity functions (Atkinson et al., 1977; Banks and Salapatek, 1978), we expected a squarewave of 1-1.5 c/deg to be at threshold given the ages tested and the contrast used. Consequently, we generally presented gratings from 1-4 c/deg in half-octave steps. Two infants were tested with even higher spatial frequencies. Thus, in each case we presented gratings at least 1½ octaves higher in frequency than the expected 70% acuity threshold.

During testing the parent held the infant on their lap or over their shoulder. The parent's view of the screen was occluded by a curtain. An observer who also could not see the stimuli viewed the infant through a 1 cm peephole just to the right of the CRT. To attract the infant's attention, the observer lowered a noise-making doll to the middle of the uniformly illuminated display screen. Once he judged that the infant was fixating the midline, the observer lifted the doll from view and initiated stimulus presentation by pressing a button. Unless the infant simply did not attend to the display, the observer guessed which side the grating had appeared on based on the infant's eye and head movements. A TRS-80 minicomputer recorded the observer's response and provided feedback. A trial was terminated when the observer responded which was generally 5-15 sec after stimulus onset. Once 5 trials were completed, the

---

[*] Squarewave gratings with a contrast of 0.80 were generally used but for some of the infants sinewave gratings with a fundamental of the same contrast were presented at the two highest spatial frequencies (2.8 and 4 c/deg). This was necessary because at those particular frequencies, the squarewave's higher harmonics appeared at times to "beat" with the X-axis waveform and, consequently, produced shimmer. Adult observers could not readily distinguish the squarewave and sinewave gratings at those frequencies so it is very unlikely that young infants could discriminate them at all.

minicomputer selected another spatial frequency randomly. The observer was unaware of the spatial frequency being presented in any given block of trials. This sequence continued until 25-35 trials were completed for all spatial frequencies[†]. The entire experiment generally required two 45-min sessions. One infant, however, was tested more extensively (see Fig. 2).

RESULTS AND DISCUSSION

The most important question to consider is whether our results exhibit the negative preference dip reported by Held et al. (1979). Figure 1 displays the observer's percent correct as a function of spatial frequency for 10 of the 11 infants. The number of trials per point in each of these psychometric functions is given in footnote[‡]. Recall that we intentionally did not present very low spatial frequencies. Consequently, the highest percent correct values were low in comparison to psychometric functions we have obtained in other experiments.

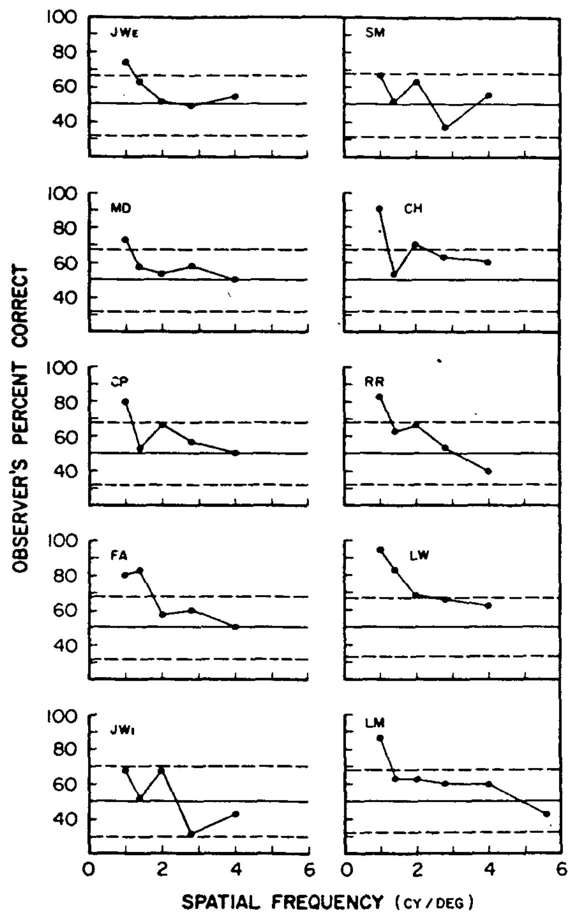

Fig. 1. Observer's percent correct as a function of spatial frequency for 10 infants. The broken lines represent performance just significantly different from 50% (*P* = 0.05, 2-tailed).

One infant was tested more extensively than the others to provide an estimate of our procedure's reliability. Figure 2 displays those data which were obtained in three different sessions in one week. The psychometric functions were consistent across this period of time.

Do our results confirm the presence of a negative preference dip? A number of points in Figs 1 and 2 were actually below 50%, but one needs to know whether these dips are just due to random variation about 50% or due to a true negative preference. The broken lines in the Figures represent the percent correct values that are just significantly different from 50% (P = 0.05, 2-tailed). The positions of the broken lines vary from infant to infant due to differences in the number of trials per point (see footnote[†]). None of the points in Figs 1 and 2 were significantly lower than 50%. (If I-tailed probabilities were used. I point would be significantly below chance.) We also calculated significance levels for the Held *et al.* data and found that 8 points were significantly below chance (P = 0.05 corresponds to 35.8%, 2-tailed). If 1-tailed probabilities were used, 11 points would be significantly below 50% (*P* = 0.05 corresponds to 38.2%). Thus, there is much clearer evidence for negative preference in their experiment than in ours.

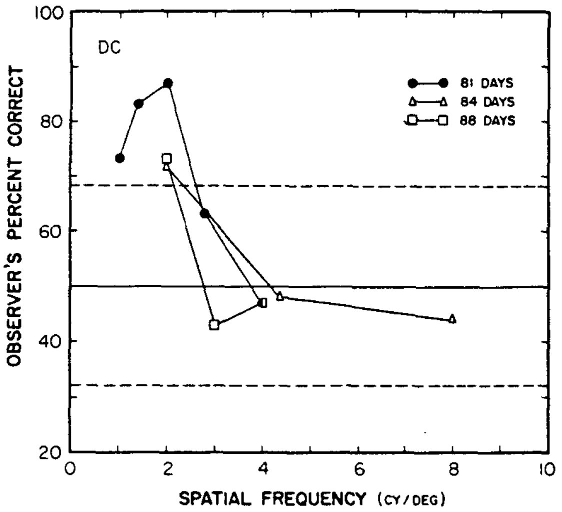

Fig. 2. Observer's percent correct as a function of spatial frequency for infant D.C. The different psychometric functions were obtained in three sessions at the ages shown.

One can also evaluate the evidence for negative preference by statistically testing for departures from randomness in the distribution of points around 50%. To do this, we noted the lowest spatial frequency at which our observer's percent correct dropped below 60% for each infant. (Note that this criterion is different than the one used in the adjoining paper; Teller *et* al., 1982). This point and all subsequent points (higher spatial frequencies) were used to create the "observed" distribution shown in Fig. 3A. Only 14 of the 34 points which met the criterion for inclusion in Fig. 3A fell at or below chance. An "expected" distribution was also constructed using a binomial sampling distribution with appropriate parameters[‡]. The two distributions did not differ significantly (X^2 = 8.53, d.f. = 11, n.s.) indicating that the tails of our psychometric functions did not deviate significantly from chance performance. We also applied this analysis to the data of Held *et al.* Figure 3B shows the corresponding "observed" and "expected" distributions. These two distributions differed significantly Q' = 30.00. d.f. = 10, p < 0.001). Inspection of Fig. 3B reveals the direction of this difference: the observed distribution had more low percent correct values than expected. Thus, different approaches, ours and that of Teller *et al.* (1982), yield the same conclusion: a statistically significant negative preference is present in the Held *et al.* data.

Our results arc clearly different from those of Held *et al.* (1979); they consistently observed a substantial negative preference dip and we did not. We will first consider what aspects of the experimental procedures might have caused the dip to appear in one case and not the other. Perhaps we did not examine the appropriate region of the psychometric function with enough precision to observe the dip. This seems unlikely because our choices of the range of spatial frequencies and step sizes between frequencies were based on the Held *et al.* experiment. The specific frequencies we presented were not the same as those in the Held *et al.* paper, but we used their approach of presenting frequencies well above the traditionally defined acuity threshold (70% point on the psychometric function) in ½-octave steps.

Perhaps the critical difference involves the observers' tasks in the two experiments. In the Held *et al.* experiment observers

---

[†] Of the 13 psychometric functions we measured, 10 were obtained with 30 trials per spatial frequency, 2 with 35 trials (L.W. and J.We. in Fig. 1), and 1 with 25 trials (J.Wi. in Fig. 1).

[‡] The parameters of the binomial sampling distribution were *n* = 30 and *P* = 0.5. That value for *n* was used because 10 of the 13 psychometric functions were obtained with 30 trials per point. If an *n* of 25 or 35 is assumed the statistical conclusions do not differ.

judged the side the infants preferred to fixate. The observers were not given feedback. Our FPL observers guessed the position of the grating based on the infants' behavior. Since feedback was given on every trial, the observers could learn to use any cue, even avoidance of one stimulus position, to maintain above chance performance. Thus we may have observed chance or above chance performance in a region where negative preference existed. Unfortunately, this argument seems unlikely because the presentation order of spatial frequencies was random and there were only five trials per block. Thus, observers would have had to change strategy quickly and appropriately to maintain high performance in regions of negative and positive preference. Our observers were not aware of any such dramatic shift in strategy.

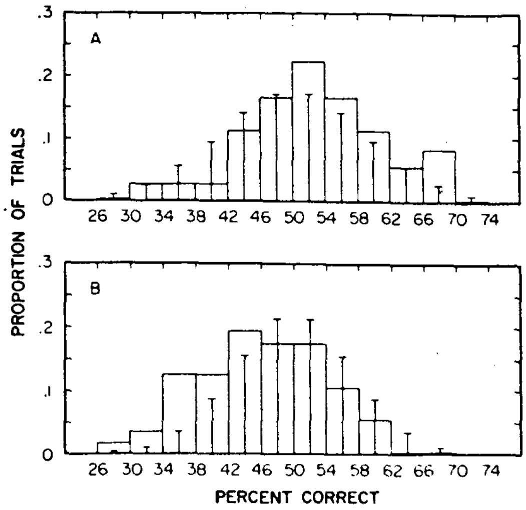

Fig. 3. Observed and expected performance distributions for points near 50% (see text for explanation). The observed distributions are represented by open bars and the expected distribution by bracketed lines. The present experiment and the Held *et al.* (1979) experiment are represented in A and B respectively. Expected distributions were generated using binomial sampling with the appropriate parameters.

Of course, there were other differences between our method and Held's. Their stimulus fields were smaller than ours. Their fields were separated widely whereas ours were adjacent to one another. The contrasts of their gratings were 13-19% higher than the contrast of ours. Any of these dissimilarities could have caused the disagreement. Unfortunately, without a clear hypothesis of what causes the negative preference dip, one cannot determine which of any number of procedural differences were influential.

Held *et al.* provided an hypothesis which states that high frequency gratings may be difficult for infants to "bring into focus thus causing accommodative stress". Consequently, the grating may become aversive. They stated that pilot data had supported this hypothesis (without indicating the nature of those data). Within the framework of this specific hypothesis, it is not clear why our procedure would have decreased the magnitude of any negative preference. If anything, lower contrast gratings, such as we used, should be more difficult to bring into focus (Banks, Dannemiller and Weaver, in preparation). Moreover, since adjacent fields present stimuli within the central visual field, an avoidance reaction due to accommodative difficulty should occur earlier in a given trial than if the fields were separated widely. Thus, it seems unlikely that "accommodative stress" would have caused avoidance in their situation and not in ours.

Another potential cause for the disagreement was raised in the preceding paper by Teller *et al.* Brightness differences between the grating and blank field might have produced a preference for the blank field once the grating was undetectable. Presumably, the blank field would have to be brighter since infants usually fixate the brighter of two otherwise identical fields (Hershenson, 1964). Teller *et al.* mention two potential sources of such a brightness difference: (1) the space-average luminance of the two stimuli may have been unequal and (2) the luminances may have been equal but perceived brightness different due to non-linear processing of the retinal image. We agree with Teller *et* al. that the first explanation seems unlikely; it does not account for the fact that Held's psychometric functions returned to chance at spatial frequencies above those where the dip was observed. We believe, in contrast to Teller *et al.*, that the second explanation is also unlikely, if non-linear processing of the grating in Held's experiment caused a reduction in its perceived brightness relative to the blank field, the same should have obtained in our experiment. In fact, since our fields were not separated spatially, one might expect simultaneous brightness contrast (if it is operative in infants) to yield a more salient brightness difference than for Held's fields which were separated by a dark surround.

In conclusion, it is not clear why Held's procedure produced the dip and ours did not. Further experimentation would be necessary to reveal the determining factors. We have shown, nonetheless, that the negative preference dip does not appear to be a robust and reliable phenomenon since it is not observed using a procedure similar to that of Held *et al.* Two implications of this are worth noting. First, Held's assertion that most preferential looking experiments have underestimated infant acuity because they ignored negative preference seems less credible. Second, considerable caution is required before using psychophysical procedures that depend on the presence of a negative preference dip. Since the factors which cause the dip are unknown, one cannot specify *a priori* for which subject populations, stimulus conditions, and response measures the dip will be present. Thus, the general utility of the Gwiazda *et al.* fast psychophysical procedure should be seriously questioned since its efficiency depends on the presence of a negative preference dip.

COMPUTER SIMULATIONS

Held and his colleagues often use the so-called "fast" preferential looking procedure to measure infant acuity (e.g. Gwiazda *et al.,* 1980a, Mohindra *et al.,* 1979; Thomas *et al.,* 1979). Gwiazda *et al.* (1980b) have recently described the procedure. Its unique features are the manner in which stimulus level is varied from trial to trial and the rules for termination. The procedure involves presenting a series of gratings in blocks of three. The blocks are ordered by spatial frequency. Step size between blocks is ½ octave. The first frequency presented is lower than the expected threshold value. Whenever the observer responds correctly by choosing the side the crating was on, the procedure advances to the next stimulus in the ordered set. Whenever an incorrect response occurs, the procedure regresses and the previous stimulus is presented. The procedure continues in this fashion until enough incorrect responses are recorded at one stimulus level to prove statistically ($P < 0.05$) that the observer is responding below 70% correct. At this point the staircase is terminated and the 70% threshold value is taken to be the spatial frequency that is ½ octave lower than the termination level.

Efficient psychophysical procedures (that is, procedures which estimate thresholds reasonably accurately within a moderate number of trials) are essential to testing infants, particularly in clinical settings, because infants generally do not tolerate lengthy sessions. The method of constant stimuli has been widely used in research settings but it requires a large number of trials to estimate threshold accurately. Gwiazda *et al.* (1980b) claim that they can obtain an accurate acuity estimate in less than 5 min using their "fast" procedure (p. 420), so it warrants serious consideration. Gwiazda *et al.* conducted two computer simulations of their procedure in order to assess its statistical properties. In one simulation the "infant" was represented by a conventional monotonic psychometric function (see their Fig. 3). In the other, a non-monotonic function with a negative preference dip was assumed (see their Fig. 4).

We have recently conducted computer simulations of a more conventional staircase procedure – the 2-down, 1-up-procedure of Levitt (1971) – in order to compare its statistical properties to those of the Gwiazda *et al.* procedure. (We also compared its properties to those of the method of constant stimuli and found that it generally estimated threshold more accurately for a given number of trials. This comparison, however, is beyond the scope of this paper.) Our computer program used a cumulative normal distribution to represent the infant's psychometric function. A number of parameters were studied: (1) the stimulus level chosen to begin the procedure, (2) the step size between levels, (3) the number of trials before terminating the staircase, and (4) the number of trials at the beginning of the staircase which were ignored in computing threshold. The computer program conducted 500 simulated experiments for each set of parameters. The results are shown in Table 1. Only cases in which the initial 5 trials were ignored are presented here. Each cell in the Table contains two values expressed in Z-units (the Z-units are standard deviations from the underlying normal distribution of the psychometric function). The first is estimation bias which is the average estimated threshold value minus the actual stimulus value corresponding to 70% correct. Positive biases indicate that an acuity value less than the spatial frequency associated with 70% correct was obtained. The second value, the one in parentheses, is the standard deviation of the 500 threshold estimates which represents the expected variability. Note that bias and variability vary with all of the staircase parameters. Variability is the more important of the two indices of accuracy (one can correct for bias if it is known), so we will emphasize it henceforth.

In their simulation of a monotonic psychometric function, Gwiazda *et al.* used a function whose abscissa was scaled in spatial frequency units rather than the Z-units we used. The function was also slightly different in shape than the cumulative normal distribution we used. To compare our results to theirs we converted their units into ours. A curve-fitting routine found the cumulative normal function which approximated their simulated psychometric function best. The Pearson correlation between the points on their function and those on the best-fitting cumulative normal was 0.99 so the fit was excellent. This best-fitting function was then used to convert Gwiazda's spatial frequency units into Z-units. The bias obtained in their simulation was -0.20Z-units and the standard deviation was 0.70 Z-units. The average number of trials to staircase termination was 38 (Gwiazda, personal communication). Since they did not vary initial stimulus level of step size in their simulation, one does not know if better performance would be obtained with different parameters. The set of conditions in our up-down staircase simulation most similar to theirs is bold face in Table 1. The bias of the up-down staircase was about 1/3 that of the Gwiazda *et al.* procedure. The variability was 35% lower. Note that different up-down parameters would yield even lower bias and variability. Obviously increasing the number of trials to termination yields lower variability, but smaller step sizes yield better results without increasing the number of trials[*]. Two points should be made concerning the results of these simulations using monotonic psychometric functions. First, the up-down staircase procedure appears to outperform the Gwiazda et *al.* "fast" procedure when equal step sizes, initial stimulus levels, and numbers of trials are used. Since the up-down procedure is also simpler, used more commonly, and its statistical properties are better known (e.g. Rose et *al.*, 1970), it seems that it offers more advantages to infant psychophysical work. Second, the parameters of staircase procedures affect their efficiency. The Gwiazda et *al.* "fast" procedure may in fact perform better if the parameters were chosen optimally. This points to the need for further study of the procedure's statistical properties.

Table 1. Bias and variability in Z-units for up–down staircase

Initial $P(C) = 85\%$

Step size	Number of trials 18	28	38	48	100
1.0	−0.10 (0.84)	−0.13 (0.69)	−0.17 (0.62)	−0.12 (0.54)	−0.18 (0.41)
0.75	0.04 (0.68)	−0.02 (0.55)	−0.08 (0.53)	−0.06 (0.49)	−0.12 (0.37)
0.50	0.15 (0.49)	0.09 (0.48)	0.04 (0.42)	0.02 (0.39)	0.00 (0.32)
0.25	0.34 (0.32)	0.24 (0.33)	0.18 (0.31)	0.17 (0.29)	0.08 (0.25)

Initial $P(C) = 80\%$

Step size	Number of trials 18	28	38	48	100
1.0	−0.04 (0.76)	−0.10 (0.68)	−0.16 (0.59)	−0.13 (0.53)	−0.16 (0.42)
0.75	−0.01 (0.65)	−0.02 (0.58)	**−0.07** **(0.52)**	−0.07 (0.48)	−0.09 (0.36)
0.50	0.10 (0.55)	0.05 (0.45)	0.02 (0.43)	0.01 (0.39)	−0.06 (0.31)
0.25	0.23 (0.36)	0.17 (0.32)	0.12 (0.33)	0.10 (0.31)	0.03 (0.25)

Initial $P(C) = 75\%$

Step size	Number of trials 18	28	38	48	100
1.0	−0.06 (0.87)	−0.11 (0.71)	−0.15 (0.60)	−0.13 (0.53)	−0.19 (0.43)
0.75	−0.01 (0.73)	−0.06 (0.59)	−0.07 (0.53)	−0.07 (0.49)	−0.13 (0.36)
0.50	0.06 (0.58)	0.03 (0.50)	−0.01 (0.44)	0.01 (0.40)	−0.05 (0.32)
0.25	0.14 (0.35)	0.01 (0.36)	0.06 (0.33)	0.08 (0.31)	0.04 (0.26)

Gwiazda *et al.* also simulated their procedure assuming a non-monotonic psychometric function with a negative preference dip. Their results were again expressed in spatial frequency units. To convert them into Z-units we used the cumulative normal function which best fit their monotonic psychometric unction since their non-monotonic function was the same except for the negative preference dip. Only 33 trials were needed on the average to reach staircase termination. Bias was -0.13 Z-units and standard deviation was 0.48. Thus, both measures of accuracy improved. These results support their claim that their procedure's efficiency depends significantly on the negative preference dip. To facilitate comparison, we conducted a simulation of the up-down procedure with a non-monotonic psychometric function. The non-monotonic function was the same as the one Gwiazda et al. used. Again we used the same step size, initial stimulus level, and number of trials they used. Bias was 0.17 and standard deviation was 0.40. Thus, once again, the up-down staircase yielded lower variability. The two points raised in the last paragraph concerning the monotonic simulations apply to the non-monotonic simulations as well.

The reader might ask whether our emphasis on the accuracy of different psychophysical procedures is justified; perhaps all of the procedures are accurate enough to allow reasonable estimates of infant thresholds. To answer this question we have to express the simulation results in terms of visual acuity. In other words, the Z-units must be converted to some index of acuity. If one fits cumulative normal functions to actual infant psychometric functions, I octave of spatial frequency on the average corresponds to about I Z-unit (Allen, 1979; Mayer, 1980, our Figs 1 and 2). The Gwiazda *et al.* simulation of monotonic psychometric functions assumed that I octave of spatial frequency corresponded to about 1.5Z-units. The actual psychometric data in Held *et* al. (1979) indicate, however, that their psychometric functions are generally steeper than this. Thus it seems safe to assume that I Z-unit in Table I and the text above corresponds to about I octave of spatial frequency. Now

---

[*] If step size is too small (e.g. 0.25), the threshold estimate becomes quite dependent on the initial stimulus level. This is evident in Table 1 where bias increases for the smallest step size. It appears that variability is low for small step size. This is probably an artifact of the way our computer simulation was conducted. For each run of 500 simulated experiments, the parameter values chosen remained constant. Thus the initial stimulus level was fixed relative to the 70% point on the psychometric function throughout a run. In practice, the initial stimulus level would not be fixed relative to the 70% point because the 70% point would vary from subject to subject. Consequently, we would expect higher variability in practice for small step sizes than we found in our simulations.

we can ask the question: how accurately can an individual infant's acuity be estimated given the statistical properties of the different staircase procedures. This can be answered by calculating the expected confidence intervals associated with different procedures. Table 2 displays the 80%, 90% and 95% confidence intervals for the up-down procedure and the Gwiazda *et* al. fast procedure. The up-down procedure's intervals are shown for the case in bold face in Table 1 (38 trials) and for the same case with a non-monotonic psychometric function (33 trials). The Gwiazda procedure's intervals are given for monotonic (38 trials) and non-monotonic psychometric functions (33 trials). These values suggest that the confidence intervals for procedures with a small number of trials are quite large. For example, if an infant yielded an acuity of 4 c/deg in the Gwiazda *et* al. fast procedure (assuming a monotonic psychometric), we could only be 90%, confident that the actual preferential looking acuity would lie between 1.8 and 8.9 c/deg. The 90% intervals assuming a nonmonotonic function would be 2.3 and 6.9 c/deg. The same intervals for a 38-trial, up-down procedure assuming a monotonic function would be 2.2 and 7.2 c/deg; for a 33-trial, up-down procedure assuming a non-monotonic function they would be 2.5 and 6.3 c/deg.

Table 2. Confidence intervals in octaves for four procedures

Confidence interval	Up-down monotonic (SD = 0.52)	Up-down non-monotonic (SD = 0.40)	Gwiazda monotonic (SD = 0.70)	Gwiazda non-monotonic (SD = 0.48)
80%	±0.67 octaves	±0.51	±0.89	±0.62
90%	±0.85	±0.66	±1.15	±0.78
95%	±1.02	±0.78	±1.37	±0.94

These large intervals indicate that short duration staircases, even though they are more efficient than other procedures like the method of constant stimuli, are simply not capable of locating a single infant's acuity threshold very accurately. The up-down procedure appears nonetheless to offer better accuracy. It should be added that these intervals are undoubtedly underestimates of the confidence intervals one would observe in practice. The calculations were based on computer simulations of infant performance. Consequently, the important effects of fluctuations in infants' behavioral state were not included.

These findings are useful in evaluating the clinical utility of different procedures. Even under ideal conditions in which the effect of infant's state was nil, quick procedures like the Gwiazda et al. fast method, or a short up-down staircase would not allow one to detect abnormal acuity confidently unless an infant's threshold differed by 1 octave or more from the appropriate population norm. Hence where greater precision is called for, either more extensive testing (that is, more trials) or different responses measures should be used.

Acknowledgements—This research was supported by NIH research grant HD 12572 and NIMH Research Scientist Development grant MH 00318 to M.S.B. and by a University of Texas Graduate Fellowship to J.L.D. The Institute of Human Development and Family Studies at the University aided in subject recruitment. We thank Janet Weaver and Anita Mendez for their help in subject testing and Katherine Wong and Robin Ditto for secretarial assistance.

REFERENCES

Allen J. (1979) Visual acuity development in human infants up to 6 months of age. Ph.D. dissertation. University of Washington.

Atkinson J., Braddick 0. and Moar K. (1977) Development of contrast sensitivity over the first 3 months of life in the human infant. Vision *Res. 17,* 1037-1044.

Banks M. S. and Salapatek P. (1978) Acuity and contrast sensitivity in 1, 2, and 3-month-old human infants. *Invest. Ophthal. visual Sci. 17,* 361-365.

Campbell F. W. and Green D. G. (1965) Optical and retinal factors affecting visual resolution. J. *Physiol. 181,* 576-593.

Gwiazda J., Brill S., Mohindra 1. And Held R. (1980a) Preferential looking acuity in infants from two to fifty-eight weeks of age. *Am. J. Optom. physiol. Opt. 57.* 428-432.

Gwiazda J., Wolfe J., Brill S., Mohindra 1. and Held R. (1980b) Quick assessment of preferential looking acuity in infants. *Am. J. Optom. physiol. Opt. 57.* 420-427.

Held R., Gwiazda J., Brill S., Mohindra 1. and Wolfe J. (1979) Infant visual acuity is underestimated because near threshold gratings are not preferentially fixated. Vision Res. 19, 1377-1379.

Hershenson M. (1964) Visual discrimination in the human newborn. *J. comp. physiol. Psychol. 58,* 270-276.

Levitt H. (1971) Transformed up-down methods in psychoacoustics. *J. acoust. Soc. Am.* 49, 467-477.

Mayer D. L. (1980) Operant preferential looking: a new technique provides estimates of visual acuity of infants and children, Ph.D. dissertation. University of Washington.

Mohindra I., Jacobson S. G.. Thomas J. and Held R. (1979) Development of amblyopia in infants. *Trans. Opthal. Soc.* U.K. 99. 344-346.

Rose R.. Teller D. Y. and Rendleman P. (1970) Statistical properties of staircase estimators. *Percept. Psychophys. 8.* 199-204.

Teller D. Y. (1979) The forced-choice preferential looking procedure. A psychophysical technique for use with human infants. *Infant Behav. D*ev, 2. 35-153.

Teller D. Y.. Mayer D. L.. Makous W. L. and Allen J. L. (1982) Do preferential looking techniques underestimate infant visual acuity? *Vision Res.* 22. 1017-1024.