

Deficient stereopsis in the normal population revisited: why current clinical stereo tests may not be adequate

Robert F Hess, PhD, DSc^{1*}, Rebecca Dillon, MSc¹, Rifeng Ding, MSc¹, Jiawei Zhou, BSc, PhD^{2*}

1. McGill Vision Research, Dept. Ophthalmology, McGill University, Montreal, PQ, Canada;
2. School of Ophthalmology and Optometry and Eye hospital, Wenzhou Medical University, Wenzhou, Zhejiang, PR China;

Correspondence: robert.hess@mcgill.ca (RFH) and zhoujw@mail.ey.ac.cn (JZ).

Correspondences:

1. Dr. Robert F. Hess

15 McGill Vision Research, Dept. Ophthalmology, McGill University, Montreal, PQ, Canada

16 robert.hess@mcgill.ca

2. Dr. Jiawei Zhou

18 School of Ophthalmology and Optometry and Eye hospital, Wenzhou Medical
19 University, Wenzhou, Zhejiang, PR China

20 zhoujw@eye.ac.cn

Number of words: 2828

24 Number of figures: 2

29 **Abstract**

30 **Significance statement:** Applied applications for occupational screening, clinical tests should
31 assess sensitivity to the sign as well as the magnitude of disparity.

32 **Purpose:** To determine why the high incidence of stereo anomaly found using laboratory tests
33 with polarity-based increment judgements (i.e., depth sign) is not reflected in clinical
34 measurements that involve single-polarity incremental judgements (i.e., depth magnitude).

35 **Methods:** An iPod-based measurement that involved the detection of an oriented shape defined
36 by a single polarity-depth increment within a random dot display was used. A staircase
37 procedure was used to gather sufficient trials to derive a meaningful measure of variance for the
38 measurement of stereopsis over a large disparity range. Forty-five adults with normal binocular
39 vision (20 - 65 years old) and normal or corrected-to-normal (0 logMAR or better) monocular
40 vision participated in this study.

41 **Results:** Observers' stereo acuities ranged between 10 and 100 arc seconds, and were
42 normally distributed on a log scale ($p = 0.90$, 2-tailed Shapiro-Wilk test). The present results
43 using a single polarity depth increment task (i.e., depth magnitude) show a similar distribution to
44 those using a similar task using the Randot preschool stereo test on individuals between the
45 ages of 19-35 using either the 4-book test ($n = 33$) or the 3-book test ($n = 40$), but very different
46 results when the iPod test involved a polarity-based increment judgement (i.e., depth sign).

47 **Conclusions:** The present clinical stereo tests are based on magnitude judgements and are
48 unable to detect the high percentage of stereo anomalous individuals in the normal population
49 revealed using depth sign judgements.

50

51

52

53

54

55 Recently, we developed a convenient method ¹ for measuring stereopsis for the clinic using an
56 iPod, which involves depth discrimination of clouds of random dots; the subject was presented
57 with two circular stimuli in depth side-by-side, one was in front of the screen (crossed disparity),
58 the other was behind the screen (uncrossed disparity); the subject was asked to detect which of
59 the two stimuli was behind the screen (i.e., a depth sign judgement). Disparity was varied using a
60 staircase procedure until a threshold was reached. One key finding was that as many as 30% of
61 people with otherwise normal vision were over 10 times worse than their fellows. While this
62 might be surprising to clinicians because the present clinical tests do not reflect this², it is not
63 surprising to vision scientists, as there have been a number of laboratory studies that have
64 suggested this ^{1, 3-9}. Although the reason for this stereo deficiency is presently unclear, Richards
65 ^{4, 5} originally suggested that it may reflect a loss of a subset of disparity-sensitive neurons that
66 process just crossed or just uncrossed disparities. One plausible reason why the current clinical
67 tests ² have failed to detect this is because they rely exclusively on detection of depth
68 magnitude (i.e., a single polarity increment judgement) rather than depth sign. To test this, we
69 have developed a different version of our previously published iPod clinical test¹ that involves a
70 detection of depths of fixed polarity (i.e., single polarity depth increment, as the majority of
71 current clinical approaches do) rather than one that is dependent of depth polarity ¹. We show
72 that the high prevalence of stereo-deficiency within the normal population can only be revealed
73 using a task that is dependent on depth polarity.

74

75

76 **Methods**

77 **Observers.** Forty-five adults (20 - 65 years old) with normal or corrected-to-normal (0 logMAR
78 or better) monocular vision and no history of binocular dysfunction participated. Written consent
79 was obtained prior to the study, which was approved by the Institutional Review Board of McGill

80 University. The described research adhered to the tenets of the Declaration of Helsinki. Except
81 the authors, all subjects were naive to the purpose of this study. The recruitment process and
82 population demographics were essentially identical to that used previously¹.

83 **Apparatus and stimulus.** The measurement was conducted by a Mac iPod (Model No: A1367)
84 running Stereogram Test app, an in-house software for IOS devices that featured a 326 ppi
85 (pixels per inch) Retina Display. The app software was written in Objective-C using IOS
86 software development kit combined with OpenGL ES 2.0. Observers viewed the stimuli
87 dichoptically through red-green anaglyph glasses at a viewing distance of 50 cm in a normal
88 light environment.

89 The stimulus was a pacman defined by static random-dots in depth on a random dot
90 background, as shown in Figure 1A. The pacman was Gaussian-windowed to blend the edge to
91 the background and to reduce monocular cues. The pacman contained randomly positioned red
92 and green dots, which had a certain offset to generate depth perception (i.e., disparity).
93 Overlapped red and green dots (or overlapped parts of dots, determined by the size of the dots)
94 were blended into an orange color by using the blending functions provided by OpenGL ES to
95 provide sub-pixel resolution.

96 The offsets between red and green dots were equal thus the pacman was perceived as in front
97 of the screen plane defined by the background random dots. In each trial, observers' task was
98 swipe the screen in the direction the pacman pointed; left, right, up or down. There was no time
99 limit for responding, as the next trial came immediately after the observers' response.

100

101 **Procedure.** Observers were asked to finish a 10-trial practice before the test. After that, a 5-min
102 test session, incorporating two separate runs, was used to estimate individuals' stereo
103 thresholds. Each run was driven by a staircase procedure in which, the initial offset between red

104 and green dots was adjustable (could set up to 40 pixels, i.e., corresponding to a stereo acuity
105 of 21.79 arc minutes) and was controlled by a 2-down/1-up staircase procedure thereafter. The
106 initial step size was 50% (relative rate), which changed after the first reversal to 10% in all
107 following trials. Since all of our participants were adults with normal vision, we set the maximal
108 offset between red and green dots to 40 pixels to ensure that it was well within D_{max} (which was
109 around 50-70 arc min)¹⁰. The staircase was terminated at the fourth reversal point. The stereo
110 threshold and its standard error were then calculated based on the last three reversals
111 averaged across the two test runs (i.e., six reversals in total).

112 **Test Configuration**

113 As is shown in Figure 1c, the following configurable parameters were provided in the
114 Stereogram Test app:

115 1) Visual acuity of the worse eye;
116 2) AMB/ANISO Side and Initial Ratio: In case of spectacle-corrected anisometropia (with or
117 without amblyopia), an image size-scaling feature was implemented to account for any
118 aniseikonia. A second program was incorporated within this app to test the degree of
119 aniseikonia; the size of the pixels could be scaled in front of the more emmetropic eye during
120 the test to eliminate any potential impact of aniseikonia. This was a part of the original clinical
121 iPod test¹ and was designed for testing anisometropic amblyopes or subjects with a high
122 degree of anisometropia. It was not utilized in this study because none of our subjects had a
123 high degree of anisometropia (i.e., > 3 dioptres). Thus, the initial ratio was set to 1:1.

124 3) Visual Alignment: In case of ocular misalignment (e.g., strabismus), an alignment calibration
125 feature was implemented to allow fusion of the two eyes' images. During the alignment, two
126 half-cross (one in red and the other one in green) were dichoptically presented to the two
127 eyes. Observers were asked to align the two half-cross into a perfect whole cross. The
128 degree to which the alignment was stable from run to run can be then assessed, as the

129 alignment offset is provided to the examiner. This was part of the original iPod clinical stereo
130 test¹ and designed for strabismic patients. It was not utilized in this study because none of
131 our participants had an ocular misalignment. Thus, x and y offsets were set to 0 px.
132 4) Repeat times: Participants were allowed to repeat individual test runs in a test session if the
133 staircase results were clearly of an anomalous form indicative of a poor determination as the
134 result of, for example, an early response mistake (finger error).

135 **Result Transformation**

136 As is shown in Figure 1d, upon completion of the test, a plot of disparity as a function of trial
137 number is provided for each test run. The disparity was recorded in pixels during the
138 measurement and was converted into minute of arc by using the following equation:

$$139 \quad Disparity = \frac{\tan^{-1}\left(\frac{n \times W_{pixel}}{D}\right)}{\pi} \times 180 \times 60 \text{arc min}$$

140 where n is the offsets between the red and green dots (in pixels), W_{pixel} is the physical width of a
141 pixel on the display and D is the distance between the subject's eyes and device's display. In
142 our study, W_{pixel} was 0.0792 mm and D was 500 mm.

143

144 **Results**

145 Stereo acuity results for 45 normal observers are plotted in Figure 2a. The results using a
146 single-polarity increment detection version of the test (Pacman) are normally distributed on this
147 log plot between 10 and 100 arc seconds ($P = .90$, 2-tailed Shapiro-Wilk test). The present
148 results show a similar distribution (on this log axis) to those of a previous study using the
149 Randot preschool stereo test, which also involves detection of the shape defined by a single
150 polarity depth increment, on individuals between the ages of 19-35 using either the 4-book test

151 (n = 33) or the 3-book test (n = 40) in Figure 2b². In Figure 2c the present results are compared
152 to those obtained previously using the iPod stereo test in which the incremental depth
153 judgement that was depended on its polarity (i.e., infront or behind)¹. The present results that
154 do not depend on depth polarity are much more tightly distributed than the previous ones¹ that
155 did depend on depth polarity, indicating much less variability across the population when depth
156 polarity is not involved.

157

158

159 **Discussion**

160 Our results confirm that the detection of stereo-deficiencies within the normal population
161 crucially depends on whether the measurement of stereopsis is dependent on the polarity of
162 depth (i.e., a polarity-based increment) or not (a single polarity increment), confirming the earlier
163 suggestions of several researchers^{1, 3-9}. As a consequence, it is a great pity that there is not
164 currently a clinical test that can detect these deficiencies in the normal population and do it in a
165 reliable way with an associated measure of variability. The current book tests (Randot and TNO
166 but see¹¹ for complete list) only measure depth detection for a fixed polarity (i.e., depth
167 magnitude) and hence would miss the stereo-deficiencies reported here. The Frisby test
168 (Burnell Co) does involve a depth polarity judgement but lacks an associated measure of
169 variance, necessary for within-subject comparisons. It also has associated monocular cues¹².
170 The current iPod stereo test offers a number of important advantages over what is currently in
171 use.

172

173 **Alternate explanations**

174 *Could it be that the Randot test² with its sharp edges can reveal finer stereo than stimuli with
175 gaussian filtered edges¹ and that polarity has nothing to do with the discrepancy?* No, other

176 polarity-based tests with sharp edges³ highlight the same discrepancy. *Could there be an issue*
177 *with the fusion of random dot stimuli that have gaussian blurred edges*¹? No, for the same
178 reason outlined in the previous answer. Gaussian filtered edges are an effective method to
179 avoid monocular cues that can be produced by sharply defined areas of depth within random
180 dot arrays. *Could it be due to an increased positional uncertainty due to the Gaussian filtered*
181 *edges leading to an uncertainty in the shape judgement?* No, Both the 2016 (Hess et al, 2016)
182 and the present study used stimuli with gaussian filtered edges. The present study relied more
183 on the form of the disparity, as the orientation in the pacman had to be detected. However, this
184 study (depth magnitude measure), unlike the 2016 study (depth sign measure) did not show a
185 wide variation in stereo acuity of normal observers. *Could it be that when stimuli of two depth*
186 *polarities are displayed together, the visual system pools them and the problem highlighted is*
187 *not one of polarity per se but anomalous depth pooling of simultaneous presented targets of*
188 *opposite polarity?*⁹, using figural instead of random dot stimuli, showed that adults (but not
189 children) exhibit deficient depth discrimination on a supra-threshold task (for disparities above
190 288 seconds) only when opposite polarity stimuli are simultaneously presented and not when
191 single polarity stimuli are presented. This suprathreshold measure is different from our threshold
192 acuity measurements but the conclusions are broadly similar. They felt, for their suprathreshold
193 measure, that anomalous disparity pooling might be responsible for two simultaneously
194 presented depth surfaces. We don't see this as such a good candidate explanation for our
195 threshold measure because the stereo anomaly we report has been shown even when the two
196 polarity surfaces are presented successively³. The role of the spatial layout of the stimuli (i.e.,
197 stimulus configuration, 2d grouping properties and overall scale of the stimulus array) or the
198 exposure duration (down to 150ms) was shown to be unimportant in the Wilcox et al study.

199

200 **Proposed explanation**

201 We have assumed that when a subject can detect a shape defined by a crossed disparity in the
202 case of the Randot or other such clinical tests that they also have access to the fact that the
203 surface appears in front of the page. This is an unwarranted assumption. Take the example of
204 motion perception, a process often compared with stereopsis. The detection of a shape defined
205 by a subset of dots moving in a particular direction does not necessarily require a knowledge of
206 the direction of movement, it could be done simply by non-directional temporal information. A
207 lesion in extra-striate cortex can abolish “motion perception” without affecting temporal change
208 perception which can be accomplished in striate cortex^{13, 14}. Our judgement of depth magnitude
209 and its associated sign may be processed at different stages alone the cortical pathway and
210 may be subject to different limitations. This is very much along the lines of a proposal made by
211 Landers and Cormack⁸, who also reported a large percentage of stereo anomalous observers in
212 the population for determining the sign but not the magnitude of the depth. They suggested that
213 while there may be a continuous representation of disparity within the visual cortex, attentive
214 readout of disparity may be a multicomponent serial process, having the following stages;
215 starting from detection of disparate areas (local correlation), magnitude of disparity (present
216 clinical stereo tests and the stimulus used in this present study), detection on the basis of
217 disparity polarity ¹) and finally discrimination of depth polarity. We propose that the percentage
218 of normal subjects that exhibit anomalous stereopsis will simply depend on which of these
219 stages is being targeted by the test; the higher the processing stage, the greater the percentage
220 of stereo anomaly within the normal population.

221

222 **Clinical as well as real world relevance**

223 Stereopsis is measured in the clinic for a number of reasons. First and foremost, as an indicator
224 of the quality of binocular function. It is also used sometimes for occupational reasons as some
225 careers require stereoscopic abilities, for example pilots, crane operators etc. Do the
226 conclusions of this laboratory study invalidate the use of current stereo tests that are based

227 solely on incremental depth judgements of a single polarity? The answer has to be yes. In
228 normal binocular vision, it is implied that we can discriminate the “infront” as well as the “behind”
229 depth of objects. An inability to do this constitutes an impairment. The majority of current
230 stereo book tests only measure in front or crossed disparities and it is assumed that the same is
231 true for behind or uncrossed disparities. That is simply not the case in up to 30% of the
232 population. In terms of occupational needs, being able to discriminate the polarity of depth
233 (infront vs behind) is obviously crucial. The next step in this research is to translate what has
234 been shown in the laboratory using computer generated disparities to assess their real-world
235 consequences.

236

237

238

239 **References**

- 240 1. Hess RF, Ding R, Clavagnier S, et al. A Robust and Reliable Test to Measure Stereopsis in
241 the Clinic. *Invest Ophthalmol Vis Sci* 2016;57:798-804.
- 242 2. Birch E, Williams C, Drover J, et al. Randot Preschool Stereoacuity Test: Normative Data
243 and Validity. *Journal of AAPOS : the official publication of the American Association for Pediatric
244 Ophthalmology and Strabismus / American Association for Pediatric Ophthalmology and
245 Strabismus* 2008;12:23-6.
- 246 3. Hess RF, To L, Zhou J, et al. Stereo Vision: The Haves and Have-Nots. *i-Perception*
247 2015;6:1-5.
- 248 4. Richards W. Anomalous Stereoscopic Depth Perception. *J Opt Soc Am* 1971;61:410-4.
- 249 5. Richards W. Stereopsis and Stereoblindness. *Experimental Brain Research* 1970;10:380-
250 8.
- 251 6. van Ee R, Richards W. A Planar and a Volumetric Test for Stereoanomaly. *Perception*
252 2002;31:51-64.
- 253 7. Dorman R, van Ee R. 50 Years of Stereoblindness: Reconciliation of a Continuum of
254 Disparity Detectors with Blindness for Disparity in near or Far Depth. *i-Perception*
255 2017;8:2041669517738542.
- 256 8. Landers DD, Cormack LK. Asymmetries and Errors in Perception of Depth from Disparity
257 Suggest a Multicomponent Model of Disparity Processing. *Perception & Psychophysics*
258 1997;59:219-31.

259 9. Wilcox LM, Hartle B, Solski A, et al. Disparity Configuration Influences Depth
260 Discrimination in Naïve Adults, but Not in Children. *Vision Res* 2017;131:106-19.
261 10. Hess RF, Liu CH, Wang Y-Z. Luminance Spatial Scale and Local Stereo-Sensitivity. *Vision*
262 *Res* 2002;42:331-42.
263 11. Read JCA. Stereo Vision and Strabismus. *Eye* 2015;29:214-24.
264 12. Holmes JM, Fawcett SL. Testing Distance Stereoacuity with the Frisby-Davis 2 (Fd2) Test.
265 *Am J Ophthalmol* 2005;139:193-5.
266 13. Hess RH, Baker CL, Jr., Zihl J. The "Motion-Blind" Patient: Low-Level Spatial and
267 Temporal Filters. *J Neurosci* 1989;9:1628-40.
268 14. Baker CL, Jr., Hess RF, Zihl J. Residual Motion Perception in a "Motion-Blind" Patient,
269 Assessed with Limited-Lifetime Random Dot Stimuli. *J Neurosci* 1991;11:454-61.
270

271

272 **Figure legends**

273 **Figure 1. An illustration of the Stereogram Test App.**

274 (a) The random-dot stimuli; (b) a screenshot of the menus in this App; (c) the test configurations
275 in the test- no aniso scaling or alignment adjustment was required in this study; (d) a plot of
276 results after finishing the test; (e) a summary of the test results.

277 **Figure 2. Histograms to show the stereo acuity results.**

278 (a) Stereo acuity results for 45 normal observers using a single polarity depth increment task
279 (Pacman shape test); (b) The present results are compared with those of a previous study
280 (shape defined by a single polarity depth increment) using the Randot preschool stereo test on
281 individuals between the ages of 19-35 using the 4-book test (blue histograms) or the 3-book test
282 (red histograms)²; (c) the present results are compared to those obtained previously using the
283 same approach but where the judgment involved stereo polarity (i.e., polarity-based increment
284 task; grey histograms)¹.

285

286

287

288

289

290

a)

b) **Stereogram Test**

21:05

Practice Mode

Start a Practice

New Test

Start a New Test

Start Using Scale Test Result

Test History

View Test History

c) **New Test**

21:06

Stereogram Test

Object Info

Object ID

Visual Info

LogMAR

AMB/ANISO Side

Initial Ratio (norm:amb)

Visual Alignment

X Axis Offset

Y Axis Offset

Start Alignment Calibration

Test Settings

Repeat Times

Initial Disparity

Distance to the Screen

d) **Scatter Plots**

Results

Valid Result (will be saved):

Average Disparity: 0.34 min

Test NO.1

Disparity in log(min of arc)

