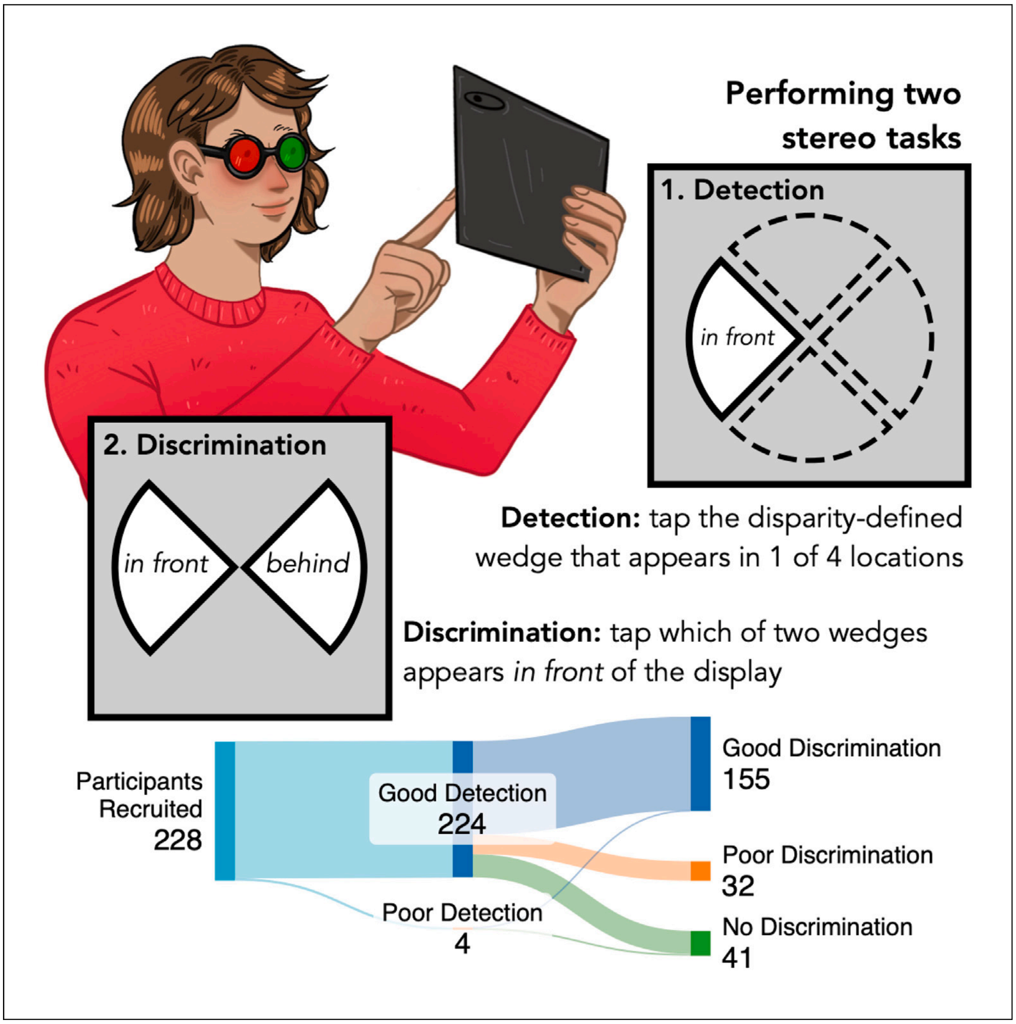
iScience

CellPress
OPEN ACCESS

Article

Stereo-anomaly is found more frequently in tasks that require discrimination between depths



Alex S. Baldwin, Seung Hyun Min, Sara Alarcon Carrillo, Zili Wang, Ziyun Cheng, Jiawei Zhou, Robert F. Hess

alexander.baldwin@mcgill.ca

Highlights

In humans, stereo-anomaly for detecting regions of disparity is quite rare (2%)

For discriminating direction of disparity, stereo-anomaly is more common (31%)

Therefore, stereo-anomaly is often due to a defect specific to identifying direction

Baldwin et al., iScience *27*, 109879
June 21, 2024
https://doi.org/10.1016/j.isci.2024.109879

iScience

CellPress
OPEN ACCESS

Article

Stereo-anomaly is found more frequently in tasks that require discrimination between depths

Alex S. Baldwin,[1,3,*] Seung Hyun Min,[1] Sara Alarcon Carrillo,[1] Zili Wang,[2] Ziyun Cheng,[2] Jiawei Zhou,[2] and Robert F. Hess[1]

SUMMARY

Within the population of humans with otherwise normal vision, there exists some proportion whose ability to perceive depth from binocular disparity is poor or absent. The prevalence of this "stereo-anomaly" has been reported to be as small as 2%, or as great as 30%. We set out to investigate this discrepancy. We used a digital tool to measure stereoacuity in tasks requiring either the detection of disparity or the discrimination of the direction of disparity. In a cohort of 228 participants, we found that 98% were able to consistently perform the detection task. Of these, only 69% consistently performed the discrimination task. The 31% of participants who had difficulty with the discrimination task could further be divided into 17% who were consistently unable to perform the task and 14% who showed limited ability. This suggests that identification of the direction of disparity requires further processing beyond merely detecting its presence.

INTRODUCTION

In human vision, the horizontal separation of the two eyes can be exploited to determine the relative distances to objects in the outside world. Fixation at a point in 3D space converges the eyes and establishes a curved surface of zero retinal disparity (the geometric horopter). Objects closer or further away from the horopter will have retinal images projected at horizontally shifted locations in the two eyes (disparity). When the objects are closer the disparity is said to be crossed. Objects further than the horopter are said to be in uncrossed disparity. Humans can show exquisite sensitivity to binocular disparity when making judgments of the relative depth of objects.

Within the human population, some individuals lack the ability to determine depth from disparity (stereo-blindness) or else show an impairment in that ability (stereo-anomaly). In some cases, this limitation is due to an identified disease (such as amblyopia) while in other cases the ability is limited with no apparent cause. There is controversy over how common this latter condition is. Some studies have found stereo-anomaly in up to 30% of participants.[1,2] Other studies have found these cases to be much rarer, affecting as few as 1%–2% of individuals.[3–7]

More recent studies have implemented stereo testing on handheld digital tablet devices; this approach has been proposed as a convenient clinical tool. One study that used a random-dot stimulus displayed on a tablet found that about 30% of the population with otherwise normal vision were up to ten times worse in their stereo sensitivity.[8] This was surprising, as the widely used Randot clinical stereo test exhibits a very narrow distribution of stereo sensitivity in that population.[3] Although these two approaches to clinical stereo measurement are different in several ways, one possible explanation of this difference is particularly interesting: stereo deficits may be specific to tasks where the polarity of the depth must be discriminated.[9,10] These stereo-anomalies would not be found with tests such as the Randot, which rely simply on detecting shapes defined by disparity.

In the present study, we compare results from disparity detection and disparity discrimination tasks using otherwise equivalent stimuli in the same individuals (Figure 1). Our goal is to identify the mechanistic basis for the previously reported deficit found in 30% of individuals with otherwise normal vision. We find that the two tasks identify stereo-anomaly with very different prevalence. In the detection task it is quite rare (2%) whereas in the discrimination task it is relatively common (31%). Therefore, the more common stereo-anomaly affects the mechanism that identifies the direction of disparity, while the detection of disparity itself may be spared.

RESULTS

Test-retest analysis for detection and discrimination tasks

Stereoacuity thresholds were obtained from a cohort of 228 participants for both disparity detection and disparity discrimination. The test-retest agreement for the two tasks is plotted in Figure 2. The data are divided according to the number of valid measurements obtained on each task. It is immediately apparent that there are many more invalid results with the discrimination task than with the detection task. This

---

[1] McGill Vision Research, Department of Ophthalmology & Visual Sciences, McGill University, Montreal, QC, Canada
[2] The Eye Hospital, School of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou, Zhejiang, China
[3] Lead contact
*Correspondence: alexander.baldwin@mcgill.ca

https://doi.org/10.1016/j.isci.2024.109879

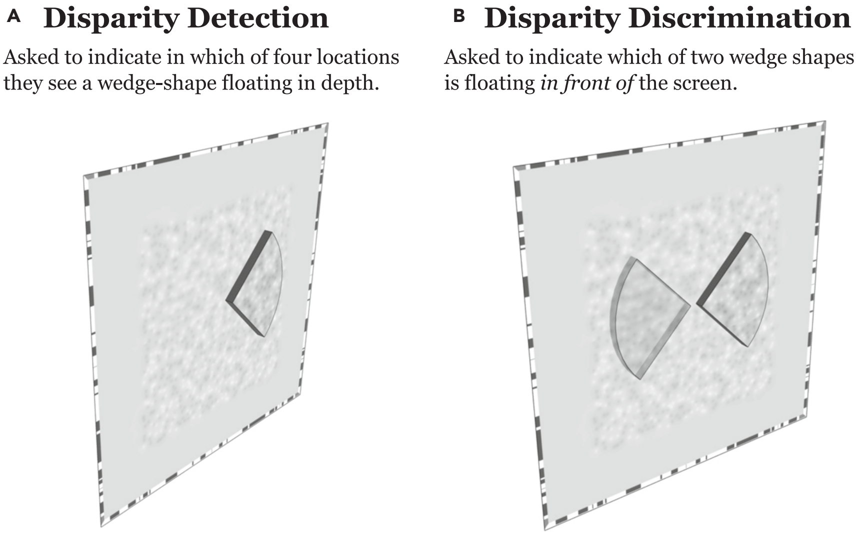


Figure 1. Diagrams showing stimulus design for the two measurements

Stimuli are rendered in a 3D projected view, with the wedges of disparity indicated. (A) shows the design for the disparity detection task, in which there were four possible target locations. This example shows the target on the right. (B) shows the design for the disparity discrimination task, in which there were two possible target locations. In this example, the wedge on the right has crossed disparity and so is the target.

indicates a specific difficulty associated with the discrimination task. The test and retest distributions for each task are visualized as histograms in Figure S1 of the supplemental information. The median discrimination thresholds are 25%–30% higher than the median detection thresholds for the test and retest measurements. The largest group obtained a valid measurement on both repetitions of both tasks (68%). The other two large groups were those who failed both repetitions (17%) or one repetition (14%) of the discrimination task. The proportion of invalid results found using the discrimination task decreases from 29% on the first ("test") measurement to 21% on the second ("retest") measurement.

Test-retest agreement was assessed using two-way mixed-effects intraclass correlation coefficient (ICC) scores. For the detection task all ICC scores fell in the "excellent" range. For discrimination, the ICC score was at the top of the "good" range.[11] These results show robust agreement between test and retest for each group of participants. A Bland-Altman analysis (Figure S2) revealed no significant bias. This means we do not find that thresholds generally improve between the test and retest measurements.

We performed an additional analysis exploring the overall proportion-correct performance of the participants in the discrimination task, by collapsing over the different disparity levels (a method we previously used to look for residual stereopsis in amblyopic participants in Alarcon Carrillo et al.[12]). We calculated the 99% binomial confidence intervals of their overall proportion correct. These were used to determine whether each participant appeared to be guessing (not significantly different from the 50% guess rate) or if there was evidence that they performed better or worse than guessing (Figures S3-S4). We found that participants who failed both test and retest measurements were typically at chance performance. On the other hand, those who failed only one repetition tended to have performed above chance in both repetitions. We can take this as evidence that the single valid result obtained from these individuals reflects an actual ability to perform the discrimination task and is not due to chance guessing.

Comparison between detection and discrimination task results

For clarity, the analysis from this point will not feature data from the small number of participants who failed any repetition of the detection task (four participants in total). We will return to these participants in our discussion. The remaining participants are those from whom valid measurements were obtained in both repetitions of the detection task. We compared the performance of each participant on the two tasks. This is presented in Figure 3. The cohort was divided into three categories: those from whom we obtained a valid discrimination threshold on both the test and retest repetitions (pink, 154 participants), those who gave two invalid discrimination results (blue, 38 participants), and those from whom we obtained a single valid discrimination result (cyan, 32 participants). We used the mean of the test and retest measurements for participants where these were both valid. For participants where only one of the two was valid, we used the value from that valid measurement only.

Among those who could perform both tasks (valid test and retest measurements for both detection and discrimination), we found a highly significant Pearson correlation between their detection and discrimination thresholds on double-log axes ($R = 0.61$, $p < 0.001$). The equation is shown in the bottom-right of Figure 3. The slope of the relationship was almost linear (0.91 on our double-log axes). We performed an additional analysis constraining the slope to be 1, in which case the offset value is 0.65. This means that the discrimination threshold for each participant was generally around 60% higher than their detection threshold.

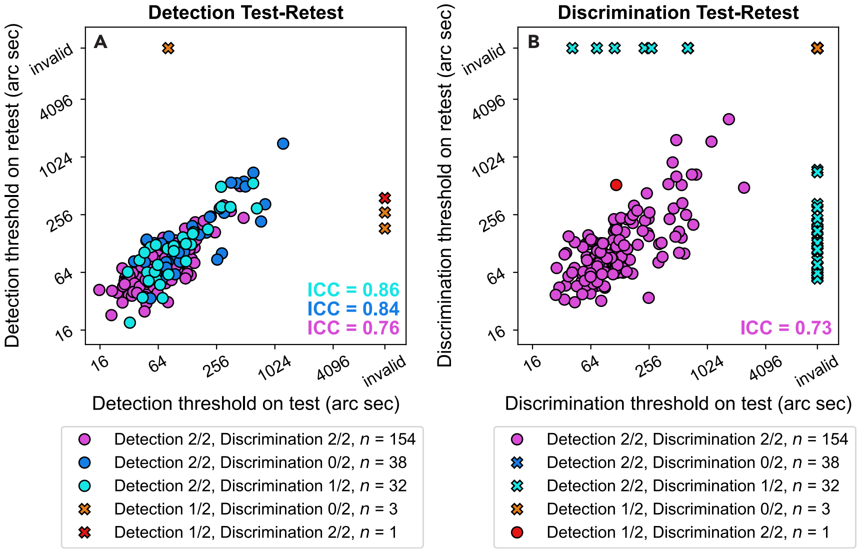


Figure 2. Test-retest scatterplots for the two tasks

Data from the detection task are shown in the left panel (A), with data from the discrimination task shown in the right panel (B). Participants are divided according to whether two, one, or zero valid measurement result was obtained in each task. Intraclass correlation coefficient (ICC) scores were calculated using a two-way mixed-effects model for the groups with at least 30 participants. We use 2/2 to refer to participants where both test and retest measurements were valid, 1/2 to refer to cases where one of the two was valid, and 0/2 where neither was valid.

The distributions of detection and discrimination thresholds from the three groups shown in Figure 3 are plotted as histograms in Figure 4. For the detection thresholds (left column), the median stereoacuity was best (69 arc sec) for those who obtained two valid results in the discrimination task. The highest median threshold (105 arc sec) was found in those who were unable to obtain a valid result in either repetition of the discrimination task. In the group who had a valid result in one of the two repetitions of the discrimination task, the median threshold was intermediate (83 arc sec). This suggests that the variation in ability to perform the discrimination task is predictive of performance in the detection task.

A Kruskal-Wallis H-test was performed on the detection threshold distributions, finding significant variation between them ($H(2) = 26.0$, $p < 0.001$). Pairwise comparisons made using Mann-Whitney U tests found a highly significant difference between detection thresholds from the groups who had either two or zero valid discrimination threshold measurements ($U = 1404$, $n_1 = 154$, $n_2 = 38$, $p < 0.001$). Analyses made with the group who had one valid discrimination threshold showed a significant difference compared to those with either two ($U = 1897$, $n_1 = 154$, $n_2 = 32$, $p = 0.041$) or zero ($U = 790$, $n_1 = 38$, $n_2 = 32$, $p = 0.032$) valid thresholds.

For the discrimination task (right column of Figure 4), the median threshold was lower in the group who obtained valid results on both repetitions (94 arc sec) compared to the group who only obtained a valid result on one repetition (133 arc sec). Comparing discrimination threshold distributions between these groups, we find again that the difference is significant ($U = 1916$, $n_1 = 154$, $n_2 = 32$, $p = 0.048$).

Finally, we compared detection thresholds against discrimination thresholds from the two groups (top and bottom rows of Figure 4). In both groups, the median discrimination threshold was higher than the detection threshold. For participants who obtained two valid measurements on both tasks, the difference in performance between the two was highly significant (Wilcoxon signed-rank test $Z = 1413$, $n = 154$, $p < 0.001$). For participants with only one valid discrimination measurement, the difference was significant ($Z = 145$, $n = 32$, $p = 0.025$).

DISCUSSION

Previous investigations of the prevalence of stereo-anomaly have found inconsistent results, ranging from 2%[4,6] to 30%.[1,2,8] These previous studies have differed in the methods used to measure stereoacuity. It has been suggested that certain tests present more difficulty (particularly to inexperienced participants) than others.[13] In this study we compared the incidence of stereoacuity measured using two tests with similar stimuli but slightly different task design in the same population.

We found that 98% of our participants gave valid results in both repetitions of a disparity detection task. The remaining 2% proportion of stereo-anomalous participants agrees with the previous studies that found lower prevalence of the condition.[4,6] Within the group that *could* perform the detection task, however, only 69% were able to perform both repetitions of an otherwise equivalent discrimination task. Of the other 31%, there were 14% who *were* able to perform one of the two repetitions. The remaining 17% were unable to perform either repetition of the discrimination task. We therefore propose that the more common stereo-anomaly concerns the identification of the sign of disparity. The 30% value from the previous literature may reflect a combination of individuals who have difficulty with discrimination tasks (our 14%) and those who are simply unable (our 17%). Our results indicate that individuals who have difficulty with the discrimination task perform worse at the detection task, compared to those who can perform disparity discrimination.

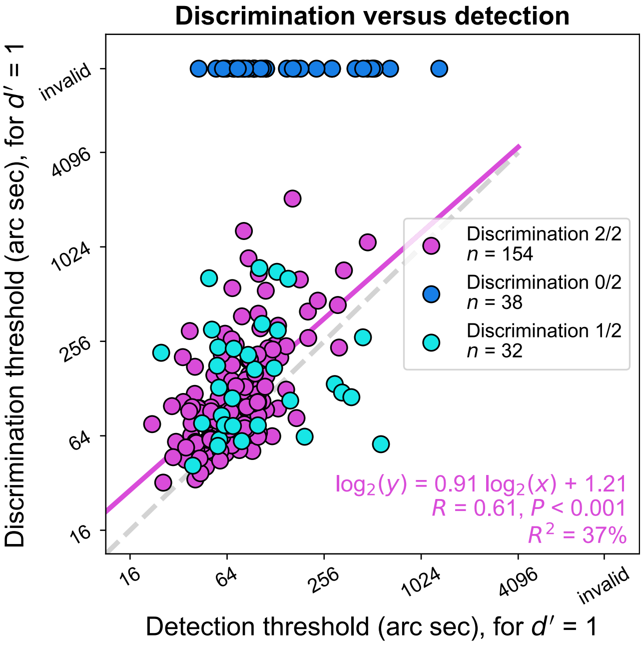


Figure 3. Scatterplot showing relationship between disparity discrimination and disparity detection performance

Thresholds obtained from both tasks are shown for the 224 participants from whom we obtained two valid detection thresholds. When thresholds were out of range the data points are placed at a disparity of 14,000 arc sec. A linear regression analysis was performed on the data from the individuals with normal stereo. This is shown by the pink line, and the statistics are reported in the bottom-right.

We performed additional analyses to investigate possible explanations for this result. Our analysis of the overall probability of responding correctly (regardless of disparity) found that the individuals we class as having difficulty with disparity discrimination tend to show above-chance performance (even when a valid result is not obtained). On the other hand, those we define as "unable" typically perform at chance when faced with a discrimination task. Although the majority of those with only one valid measurement performed better in the retest measurement compared to the test measurement, our Bland-Altman analysis did not reveal a general test-retest bias for either the detection or discrimination task.

Results from other tasks can further narrow down our interpretation. In our other studies using similar methods and stimulus design, we did not find the large proportion of stereo-anomalous individuals identified in this study. In Carrillo et al. (2020)[14] we looked for, but did not find, stereo-blindness specific to detecting stimuli in one direction of disparity.[15] We speculate that a deficit of this kind would also give a different result in the current study, as the ability to see one direction of disparity should still allow participants to perform the discrimination task. Similarly, Alarcon Carrillo et al.[12] introduced an "odd-one-out" version of the test where there were four wedges. Three wedges had the same disparity (crossed or uncrossed); the target was the fourth wedge, which had the opposite sign of disparity. Among the 17 control participants in that study, we also did not find any stereo-anomalous participants. This suggests that identifying the direction of disparity may introduce an additional difficulty for some participants which is not present in tasks which only require them to discriminate an odd one out.

Specific difficulties related to the discrimination of disparity information have been noted previously.[16] It has been suggested that sensitivity in depth discrimination tasks relates to the past stereoscopic viewing experience of the participant.[17] Recent work with experts trained to identify features in stereoscopic aerial photographs found an enhanced use of disparity information to distinguish hedges from ditches (by the direction of disparity) but no association between this ability and a clinical test of stereoacuity that did not require the identification of disparity direction.[18] Perhaps conversely, a negative association with age has also been found. In discrimination experiments using stimuli with both crossed and uncrossed targets (similar in that way to our design), better performance was found in children aged 4–6 years old compared to adults (who were at chance[19]). We expect the portable tablet-based testing approach used in the current study will provide an efficient means of identifying the possible stimulus and task factors that show this stereo-anomaly in future work.

Our findings may align with recent work on the perception of depth from anti-correlated stereograms. Feedforward models of depth perception based on V1 neurophysiology predict a reversal of perceived disparity in these stimuli, but such a result is not typically found experimentally.[20,21] Recent work suggests that this predicted *mis*-perception of disparity direction is usually prevented by feedback from higher visual areas.[22,23] A defect in this feedback process may be responsible for the deficit we find in some individuals for the discrimination task.

Conclusion

We found that an inability to detect disparity-defined targets in a random-dot stimulus was quite rare (2% in our sample). Far more common was a difficulty (14%) or inability (17%) in discriminating the *direction* of disparity. We propose this finding explains why previous studies that

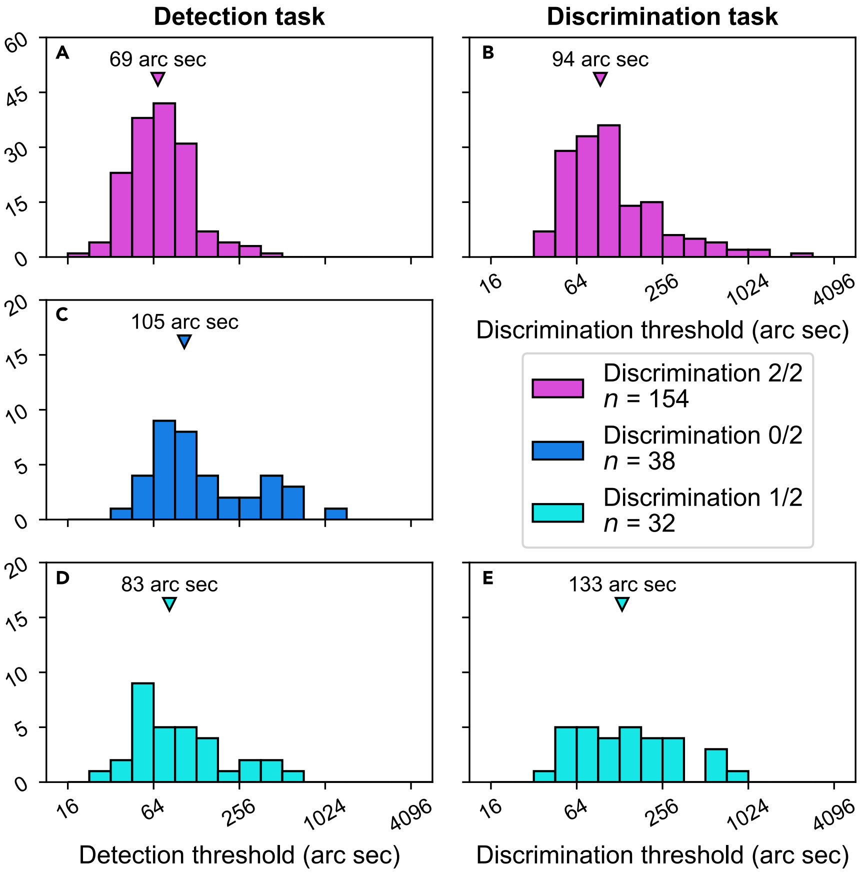


Figure 4. Histograms of thresholds obtained on the detection (first column) and discrimination (second column) tasks

The three rows show thresholds obtained from the three groups, into which they are divided by the number of valid results obtained on the discrimination task (as in Figure 3). Panels A and B show detection and discrimination performance (respectively) from the group who obtained two valid results on the discrimination task. Panel C shows the detection performance from the group who obtained no valid results on the discrimination task. Panels D and E show the detection and discrimination performance from the group who obtained only one valid result on the discrimination task. The triangle marker indicates the median of each distribution.

used different tasks to investigate the prevalence of stereo-anomaly found disparate results. The studies that found a lower prevalence could be performed using detection alone. The studies finding a higher prevalence required participants to discriminate the direction of disparity. We therefore conclude that: i) the mechanism by which depth is perceived from disparity in human vision has a component that determines the sign, which can be separately impaired even in otherwise normal vision and ii) a comprehensive assessment of stereo-ability should measure both detection and discrimination performance.

Limitations of the study

We did not set out to recruit a diverse group of participants for this study. Our participant cohort was 66% female, and the study was not designed to perform any analyses investigating sex as a factor. There have been relatively few studies looking at sex-related differences in visual perception; however research investigating other (non-stereo) tasks has found evidence of such differences.[24] Additionally, the participants were recruited and tested at one site (Wenzhou Medical University) and were of a relatively narrow age range. Performance on stereo tasks is expected to change over the lifespan,[4] and so the proportions we report may also have been different had we included older or younger participants in our sample. It is notable however that the prevalences we do find are in line with those reported by previous studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- ● EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- ● METHOD DETAILS
  - ○ Apparatus
  - ○ Stimuli
  - ○ Procedure
- ● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109879.

ACKNOWLEDGMENTS

The authors thank Ran Zhang for assisting in data collection. For our graphical abstract we would like to thank Cinzia Sekeres and SankeyMATIC (sankeymatic.com). This work was supported by the Natural Sciences and Engineering Research Council of Canada (#2016-03740 to R.F.H.) and the Natural Science Foundation for Distinguished Young Scholars of Zhejiang Province, China (#LR22H120001 to J.Z.).

AUTHOR CONTRIBUTIONS

A.S.B.: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review and editing, visualization, supervision, funding acquisition; S.H.M.: investigation, writing – review and editing; S.A.C.: formal analysis, writing – review and editing; Z.W.: investigation, writing – review and editing; Z.C.: investigation, writing – review and editing; J.Z.: resources, writing – review and editing, supervision, project administration, funding acquisition; R.F.H.: conceptualization, methodology, resources, writing – review and editing, supervision, project administration, funding acquisition.

DECLARATION OF INTERESTS

A.S.B. and R.F.H. are both inventors on patent(s) and other intellectual properties (including Hess and Baldwin, 2020) concerning stereovision and the measurement and treatment of disorders of binocular vision such as amblyopia. Some of these technologies have been commercially licensed by McGill University to Novartis International AG. The inventors have (separate from the work described here) worked to develop these technologies for clinical use through a research agreement involving Novartis and the Research Institute of the McGill University Health Centre.

Received: December 6, 2023
Revised: February 28, 2024
Accepted: April 29, 2024
Published: May 3, 2024

REFERENCES

1. Richards, W. (1970). Stereopsis and stereoblindness. Exp. Brain Res. *10*, 380–388. https://doi.org/10.1007/BF02324765.
2. Hess, R.F., To, L., Zhou, J., Wang, G., and Cooperstock, J.R. (2015). Stereo vision: the haves and have-nots. Percept *6*, 2041669515593028. https://doi.org/10.1177/2041669515593028.
3. Birch, E., Williams, C., Drover, J., Fu, V., Cheng, C., Northstone, K., Courage, M., and Adams, R. (2008). Randot® Preschool Stereoacuity Test: Normative data and validity. J. Am. Assoc. Pediatr. Ophthalmol. Strabismus *12*, 23–26. https://doi.org/10.1016/j.jaapos.2007.06.003.
4. Bohr, I., and Read, J.C.A. (2013). Stereoacuity with Frisby and revised FD2 stereo tests. PLoS One *8*, e82999. https://doi.org/10.1371/journal.pone.0082999.
5. Coutant, B.E., and Westheimer, G. (1993). Population distribution of stereoscopic ability. Ophthalmic Physiol. Opt. *13*, 3–7. https://doi.org/10.1111/j.1475-1313.1993.tb00419.x.
6. Newhouse, M., and Uttal, W.R. (1982). Distribution of stereoanomalies in the general population. Bull. Psychon. Soc. *20*, 48–50. https://doi.org/10.3758/BF03334799.
7. Zaroff, C.M., Knutelska, M., and Frumkes, T.E. (2003). Variation in stereoacuity: normative description, fixation disparity, and the roles of aging and gender. Invest. Ophthalmol. Vis. Sci. *44*, 891–900. https://doi.org/10.1167/iovs.02-0361.
8. Hess, R.F., Ding, R., Clavagnier, S., Liu, C., Guo, C., Viner, C., Barrett, B.T., Radia, K., and Zhou, J. (2016). A Robust and Reliable Test to Measure Stereopsis in the Clinic. Invest. Ophthalmol. Vis. Sci. *57*, 798–804. https://doi.org/10.1167/iovs.15-18690.
9. Hess, R.F., Dillon, R., Ding, R., and Zhou, J. (2019). Deficient stereopsis in the normal population revisited: why current clinical stereo tests may not be adequate. bioRxiv *585976*, 1–14. https://doi.org/10.1101/585976.
10. Landers, D.D., and Cormack, L.K. (1997). Asymmetries and errors in perception of depth from disparity suggest a multicomponent model of disparity processing. Percept. Psychophys. *59*, 219–231. https://doi.org/10.3758/BF03211890.
11. Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. *6*, 284–290. https://doi.org/10.1037/1040-3590.6.4.284.
12. Alarcon Carrillo, S., Hess, R.F., Mao, Y., Zhou, J., and Baldwin, A.S. (2023). Amblyopic stereo vision is efficient but noisy. Vision Res. *210*, 108267. https://doi.org/10.1016/j.visres.2023.108267.
13. Bosten, J.M., Goodbourn, P.T., Lawrance-Owen, A.J., Bargary, G., Hogg, R.E., and Mollon, J.D. (2015). A population study of

binocular function. Vision Res. *110*, 34–50. https://doi.org/10.1016/j.visres.2015.02.017.
14. Alarcon Carrillo, S., Baldwin, A.S., and Hess, R.F. (2020). Factors limiting sensitivity to binocular disparity in human vision: Evidence from a noise-masking approach. J. Vis. *20*, 9. https://doi.org/10.1167/jov.20.3.9.
15. Richards, W. (1971). Anomalous stereoscopic depth perception. J. Opt. Soc. Am. *61*, 410–414. https://doi.org/10.1364/JOSA.61.000410.
16. Harris, J.M., Chopin, A., Zeiner, K., and Hibbard, P.B. (2012). Perception of Relative Depth Interval: Systematic Biases in Perceived Depth. Q. J. Exp. Psychol. *65*, 73–91. https://doi.org/10.1080/17470218.2011.589520.
17. Stransky, D., Wilcox, L.M., and Allison, R.S. (2014). Effects of Long-Term Exposure on Sensitivity and Comfort with Stereoscopic Displays. ACM Trans. Appl. Percept. *11*, 1–13. https://doi.org/10.1145/2536810.
18. Skog, E., Meese, T.S., Sargent, I.M.J., Ormerod, A., and Schofield, A.J. (2024). Classification images for aerial images capture visual expertise for binocular disparity and a prior for lighting from above. J. Vis. *24*, 11. https://doi.org/10.1167/jov.24.4.11.
19. Wilcox, L.M., Hartle, B., Solski, A., Mackenzie, K.J., and Giaschi, D. (2017). Disparity configuration influences depth discrimination in naïve adults, but not in children. Vision Res. *131*, 106–119. https://doi.org/10.1016/j.visres.2016.12.010.
20. Chen, P.-Y., Chen, C.-C., and Tyler, C.W. (2021). A gain-control disparity energy model for perceived depth from disparity. Vision Res. *181*, 38–46. https://doi.org/10.1016/j.visres.2020.12.008.
21. Cumming, B.G., and Parker, A.J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. Nature *389*, 280–283. https://doi.org/10.1038/38487.
22. Zhaoping, L. (2021). Contrast-reversed binocular dot-pairs in random-dot stereograms for depth perception in central visual field: Probing the dynamics of feedforward-feedback processes in visual inference. Vision Res. *186*, 124–139. https://doi.org/10.1016/j.visres.2021.03.005.
23. Zhaoping, L., and Ackermann, J. (2018). Reversed Depth in Anticorrelated Random-Dot Stereograms and the Central-Peripheral Difference in Visual Inference. Perception *47*, 301006618758571. https://doi.org/10.1177/0301006618758571.
24. Shaqiri, A., Roinishvili, M., Grzeczkowski, L., Chkonia, E., Pilz, K., Mohr, C., Brand, A., Kunchulia, M., and Herzog, M.H. (2018). Sex-related differences in vision are heterogeneous. Sci. Rep. *8*, 7521. https://doi.org/10.1038/s41598-018-25298-8.
25. Hess, R.F., and Baldwin, A. (2020). A.S. System and method for digital measurement of stereo vision. International Application Published Under the Patent Cooperation Treaty. WO 2020/146952 A1.
26. Atchison, D.A., Lee, J., Lu, J., Webber, A.L., Hess, R.F., Baldwin, A.S., and Schmid, K.L. (2020). Effects of simulated anisometropia and aniseikonia on stereopsis. Ophthalmic Physiol. Opt. *40*, 323–332. https://doi.org/10.1111/opo.12680.
27. Tittes, J., Baldwin, A.S., Hess, R.F., Cirina, L., Wenner, Y., Kuhli-Hattenbach, C., Ackermann, H., Kohnen, T., and Fronius, M. (2019). Assessment of stereovision with digital testing in adults and children with normal and impaired binocularity. Vision Res. *164*, 69–82.
28. Webber, A.L., Schmid, K.L., Baldwin, A.S., and Hess, R.F. (2020). Suppression rather than visual acuity loss limits stereoacuity in amblyopia. Invest. Ophthalmol. Vis. Sci. *61*, 50. https://doi.org/10.1167/iovs.61.6.50.
29. Webber, A.L., Wood, J.M., Thompson, B., and Birch, E.E. (2019). From suppression to stereoacuity: a composite binocular function score for clinical research. Ophthalmic Physiol. Opt. *39*, 53–62. https://doi.org/10.1111/opo.12599.
30. Chopin, A., Chan, S.W., Guellai, B., Bavelier, D., and Levi, D.M. (2019). Binocular non-stereoscopic cues can deceive clinical tests of stereopsis. Sci. Rep. *9*, 5789. https://doi.org/10.1038/s41598-019-42149-2.
31. O'shea, R.P., and Blake, R. (1987). Depth without disparity in random-dot stereograms. Percept. Psychophys. *42*, 205–214. https://doi.org/10.3758/BF03203072.
32. Serrano-Pedraza, I., Vancleef, K., and Read, J.C.A. (2016). Avoiding monocular artifacts in clinical stereotests presented on column-interleaved digital stereoscopic displays. J. Vis. *16*, 13. https://doi.org/10.1167/16.14.13.
33. Prins, N., and Kingdom, F.A.A. (2018). Applying the Model-Comparison Approach to Test Specific Research Hypotheses in Psychophysical Research Using the Palamedes Toolbox. Front. Psychol. *9*, 1250. https://doi.org/10.3389/fpsyg.2018.01250.
34. Eaton, J.W., Bateman, D., Hauberg, S., and Wehbring, R. (2020). GNU Octave version 6.1.0 manual: a high-level interactive language for numerical computations.
35. Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. *9*, 90–95. https://doi.org/10.1109/MCSE.2007.55.
36. Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2.
37. Vallat, R. (2018). Pingouin: statistics in python. J. Open Source Softw. *3*, 1026.
38. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.
39. Seabold, S., and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In 9th python in science conference.
40. Gelman, A., and Hill, J. (2007). 4 Linear regression: before and after fitting the model. In Data analysis using regression and multilevel/hierarchical models Analytical methods for social research (Cambridge Univ. Press), pp. 53–198.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Psychophysical Data	figshare	https://doi.org/10.6084/m9.figshare.24526675
Software and algorithms		
Palamedes Toolbox v1.11.11	https://www.palamedestoolbox.org/	https://doi.org/10.3389/fpsyg.2018.01250
Matplotlib v3.7.1	https://matplotlib.org/	https://doi.org/10.1109/MCSE.2007.55
NumPy v1.25.2	https://numpy.org/	https://doi.org/10.1038/s41586-020-2649-2
SciPy v1.10.1	https://scipy.org/	https://doi.org/10.1038/s41592-019-0686-2
Pingouin v0.5.3	https://pingouin-stats.org/	https://doi.org/10.21105/joss.01026
Statsmodels v0.14.0	https://www.statsmodels.org/	https://doi.org/10.25080/Majora-92bf1922-011

RESOURCE AVAILABILITY

Lead contact

Requests for any further information and data should be sent to the lead contact, Alex Baldwin (alexander.baldwin@mcgill.ca).

Materials availability

Not applicable, no reagents were generated in this study.

Data and code availability

- Data: Psychophysical data deposited to figshare, https://doi.org/10.6084/m9.figshare.24526675. These data will be available as of the date of publication.
- Code: The analysis performed in this study uses code from open-source Python packages, and the Palamedes Toolbox referenced in the quantification and statistical analysis section and in our key resources table.
- All other items: Any additional information required to re-analyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We tested 228 participants (aged 17-36, 150 female). Participants were recruited at Wenzhou Medical University. All participants had normal vision or were corrected-to-normal with prescribed optical correction (above 20/20). Participants that reported any visual disorders (cataracts, glaucoma, or amblyopia) were excluded from the study. Written informed consent was given by all participants. All testing was performed in accordance with the Declaration of Helsinki and approved by the Research Ethics Board of the McGill University Health Centre and Wenzhou Medical University.

METHOD DETAILS

Apparatus

Testing was carried out using two stereoacuity apps developed at McGill Vision Research.[25] These are variations on the Baldwin-Hess stereo test design that has been used in a number of previous studies.[12,14,26–29] These were installed on a 5th generation Apple iPad mini (model A2133; Apple Inc, Cupertino, CA). Testing was performed at a viewing distance of 32 cm (maintained with the aid of a length of string attached to the tablet). At this distance, the display had 72 pixels per degree of visual angle. The screen brightness was set to about half the maximum level, approximately 150 cd/m^2. The iPad's automatic brightness control feature was turned off.

Stimuli

The stimuli in the detection and discrimination tasks were generated using methods similar to those reported in our previous studies.[14,12,27] These studies share the use of stimuli generated from spatially-bandpass (peak spatial frequency 0.4 c/deg, spatial frequency bandwidth ±2.2 octaves) isotropic log-Gabors which are composed into a field of "fuzzy" dots on a grey background. These dots reduce the impact of optical factors on the clarity of the stimulus image (by removing higher spatial frequencies), and also allow for fine positioning by sub-pixel interpolation without introducing regions of blur in the stimulus. The current study uses a mixture of dots with dark and light centres, after Alarcon Carrillo et al. (2020).[14] They were placed with an average dot-to-dot distance of 32 arc min. Stimuli were presented at a Michelson contrast of 80%.

Disparity was introduced into the stimuli by shifting the horizontal position of the dots. For the disparity detection task (Figure 1A) the dots falling within a wedge-shaped target zone (6 deg radius) were given a crossed disparity equal to half of the total stimulus disparity. The remaining "background" dots were given an equal amount of "uncrossed" disparity. The total stimulus disparity therefore is that between the foreground (wedge) and background dots. This has the benefit of applying the same positional shift to every dot in the stimulus, safeguarding against both binocular and monocular cues that might allow a participant without stereo vision to perform this task.[30–32]

The disparity discrimination task featured two wedge-shaped targets to which disparity was applied (Figure 1B), with the remaining dots being rendered at the plane of the display. The target wedge was rendered with crossed disparity, and the other wedge with uncrossed disparity.

Procedure

The disparity detection task was a four-alternative forced-choice task (target shown in the left, right, top, or bottom position), whereas the discrimination task was two-alternative forced-choice (target on the left or on the right). Disparity thresholds were obtained for both the detection and discrimination tasks. The disparities were selected on a trial-to-trial basis using a pair of interleaved two-down one-up staircase algorithms to sample the informative region of the psychometric function. The initial disparity was 1024 arc sec and the step size was a factor of two. Testing was performed indoors, and in a single session for each participant. The detection and discrimination tasks were performed in a semi-random order, constrained by the requirement that each participant needed to perform each type of measurement twice ("test" and "retest").

QUANTIFICATION AND STATISTICAL ANALYSIS

Psychometric functions were fitted using the Palamedes toolbox[33] in GNU Octave.[34] Further analysis and graph plotting was performed in Python (Python Software Foundation, Wilmington, DE) using the Matplotlib[35] and NumPy[36] libraries. Intraclass Correlation Coefficients (ICC) were calculated using Pingouin.[37] Mann-Whitney U tests, Kruskal-Wallis H-tests, Wilcoxon signed-rank tests, and t-tests were performed using SciPy.[38] Statsmodels[39] was used to calculate binomial confidence intervals. Details of the main statistical results can be found in the Results section, in Figures 2, 3, and 4, and in their captions. Further statistical results can be found in the supplemental information, in Figures S1-S4, and in their captions.

In all analyses, we took the $\log_2$ of the disparity values and performed our fitting and analysis with this logarithmic scale. This is a typical approach taken to stereoacuity data,[7,8] and appropriate for a measurement which must be positive and can span a large range.[40] To fit the psychometric function, we used a Logistic function and obtained the threshold at 55.20% correct for the 4-alternative forced-choice detection task and 76.02% correct for the 2-alternative forced-choice discrimination task (corresponding to a d' of 1 in both cases). An advantage of this approach was that it allowed for an estimate of the error associated with the threshold parameter, which can be used to reject data where atypical behaviour (i.e. by guessing alone) resulted in thresholds which were poorly constrained.

We obtained standard errors of the stereoacuity thresholds through non-parametric bootstrapping. We set criteria on the result of the psychometric function fitting to determine whether a valid result had been obtained. Invalid results were those with a large standard error ($>2 \log_2$ units) or those where the computed threshold was outside of the range measurable by the device (see Appendix B in Alarcon Carrillo et al., 2023). In this case that limit was 4,096 arc seconds (12 $\log_2$ units).