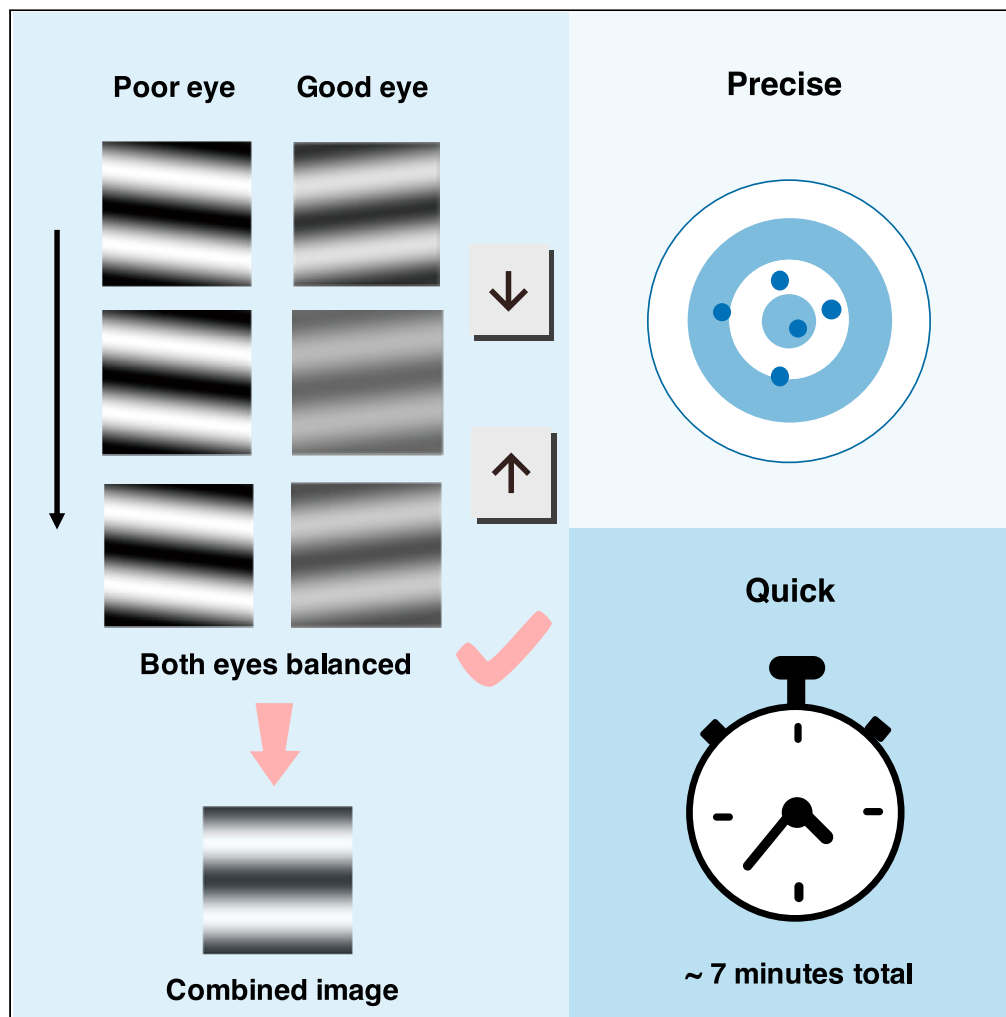


## Article

## A clinically convenient test to measure binocular balance across spatial frequency in amblyopia



Seung Hyun Min,  
Yu Mao, Shijia  
Chen, Zhifen He,  
Robert F. Hess,  
Jiawei Zhou

zhifen0821@163.com (Z.H.)  
robert.hess@mcgill.ca (R.F.H.)  
zhoujw@mail.eye.ac.cn (J.Z.)

**Highlights**

Measuring binocular  
balance of amblyopes is  
difficult and time-  
consuming (>30 min)

We introduce a  
psychophysical test that is  
reliable and quick (7 min)

## Article

## A clinically convenient test to measure binocular balance across spatial frequency in amblyopia

Seung Hyun Min,<sup>1,2,3</sup> Yu Mao,<sup>1,3</sup> Shijia Chen,<sup>1</sup> Zhifen He,<sup>1,\*</sup> Robert F. Hess,<sup>2,\*</sup> and Jiawei Zhou<sup>1,4,\*</sup>

## SUMMARY

**Amblyopia is a visual disorder that originates from the brain. It exhibits no pathology in the eye. Studies have shown that measuring both visual acuity and binocular balance for assessing amblyopia could be more helpful. However, tests that measure binocular balance are time-consuming, often exceeding 30 min. Their long test durations prevent them from being used in the clinic. For this reason, we have developed a quick (i.e., about 7 min) and precise tool that quantitatively measures binocular balance of patients with amblyopia. The new test can capture binocular imbalance that is typically exhibited at high spatial frequency in amblyopes. In addition, it has an excellent test-retest reliability and repeatability between two experimental sessions. We hope that our newly developed test can pave the road for physicians and researchers to better assess and diagnose amblyopia and other visual disorders that disrupt binocular balance beyond the laboratory.**

## INTRODUCTION

Binocular vision is essential for a normal daily life. For instance, activities such as reading (Johansson et al., 2014), hand-eye coordination (Suttle et al., 2011), driving (Adrian et al., 2019), and sports (Heinen and Vinken 2011) require a normal binocular vision. If binocular vision is perfectly normal, both eyes are balanced and contribute equally to construct a binocular image. However, when it is disrupted, one eye contributes more to a binocular percept than the other eye, resulting in an interocular imbalance. The most extreme types of the imbalance are observed in clinical conditions such as strabismus, anisometropia, and amblyopia (Hess et al., 2014; Holopigian et al., 1986; Levi et al., 1980; Zhou et al., 2013).

Both eyes combine to form a single, binocular image by a process known as binocular combination, which can be measured in the laboratory to assess interocular balance. During a test that measures binocular combination, fusible but different stimuli from each eye can be displayed at multiple contrasts (i.e., multiple visibility levels). If the eyes are perfectly balanced, the contrast in each eye resulting in a balanced percept should be identical. In this case, the interocular contrast ratio to achieve a binocular balance will be 1 (contrast of one eye/contrast of the other eye = 1). However, if the right eye requires a higher contrast than the left eye to reach a perfect interocular balance, then the left eye is said to be more dominant. In this instance, the interocular contrast ratio between the left and right eyes to reach a balance deviates from 1. Several binocular combination tests exist, such as phase (Ding and Sperling 2006; Huang et al., 2009), contrast (Ding et al., 2013a, 2013b; Huang et al., 2010), second-order modulation (Zhou et al., 2014, 2016), global motion coherence (Mansouri et al., 2008), and a global orientation coherence (Zhou et al., 2013). Although the nature of the visual stimuli varies in all these tasks, the purpose of the tasks is to measure the interocular contrast ratio where both eyes contribute equally. They have been used to study visual disorders and neural plasticity in the human visual cortex.

However, these tests that evaluate binocular combination can be limited by the spatial resolution of the screen. For this reason, they are restricted to display stimuli at a lower spatial frequency (Ding and Sperling 2006; Ding et al., 2013a, 2013b). This limitation poses a challenge for studying amblyopia, a common eye disease whose spatial deficits are mainly at the mid to high spatial frequencies (Beylerian et al., 2020; Ding et al., 2013a, 2013b; Kwon et al., 2015; Reynaud and Hess 2016). Recent studies show that the binocular imbalance in amblyopia worsens at higher spatial frequencies even if the visual stimuli are shown at a constant suprathreshold contrast (Ding et al., 2013a, 2013b; Kwon et al., 2015; Reynaud and Hess 2016). For instance,

<sup>1</sup>School of Ophthalmology and Optometry, Affiliated Eye Hospital, State Key Laboratory of Ophthalmology, Optometry and Vision Science, Wenzhou Medical University, Wenzhou, Zhejiang, China

<sup>2</sup>McGill Vision Research, Department of Ophthalmology and Visual Sciences, McGill University, Montreal, Canada

<sup>3</sup>The authors contributed equally

<sup>4</sup>Lead contact

\*Correspondence: zhifen0821@163.com (Z.H.), robert.hess@mcgill.ca (R.F.H.), zhoujw@mail.eye.ac.cn (J.Z.)  
<https://doi.org/10.1016/j.isci.2021.103652>



Ding et al. measured the binocular combination for amblyopes at three spatial frequencies (0.68, 1.36, and 2.72 c/deg) (Ding et al., 2013a, 2013b). They observed that the contrast of the dominant eye had to be lowered to achieve a perfect binocular balance at higher spatial frequencies. This finding indicates that the relative contribution of the amblyopic eye decreases as spatial frequency increases. However, this study was not able to explore spatial frequencies beyond 2.72 c/deg due to the limitation posed by the spatial resolution of the screen. To address this issue, Kwon et al. introduced a band-pass filtered letter chart, which is a binocular rivalry task (Kwon et al., 2015). In their study, a different letter was shown to each eye, inducing a visual competition between the eyes (i.e., binocular rivalry). Binocular imbalance of amblyopes was measured from 0.5 to 5 c/deg. They demonstrated that the binocular imbalance increased by 19% at higher spatial frequencies. However, binocular rivalry and combination are different processes. The former is a process of interocular conflict over time, whereas the latter is a process by which the input from both eyes combines to form a single, binocular percept.

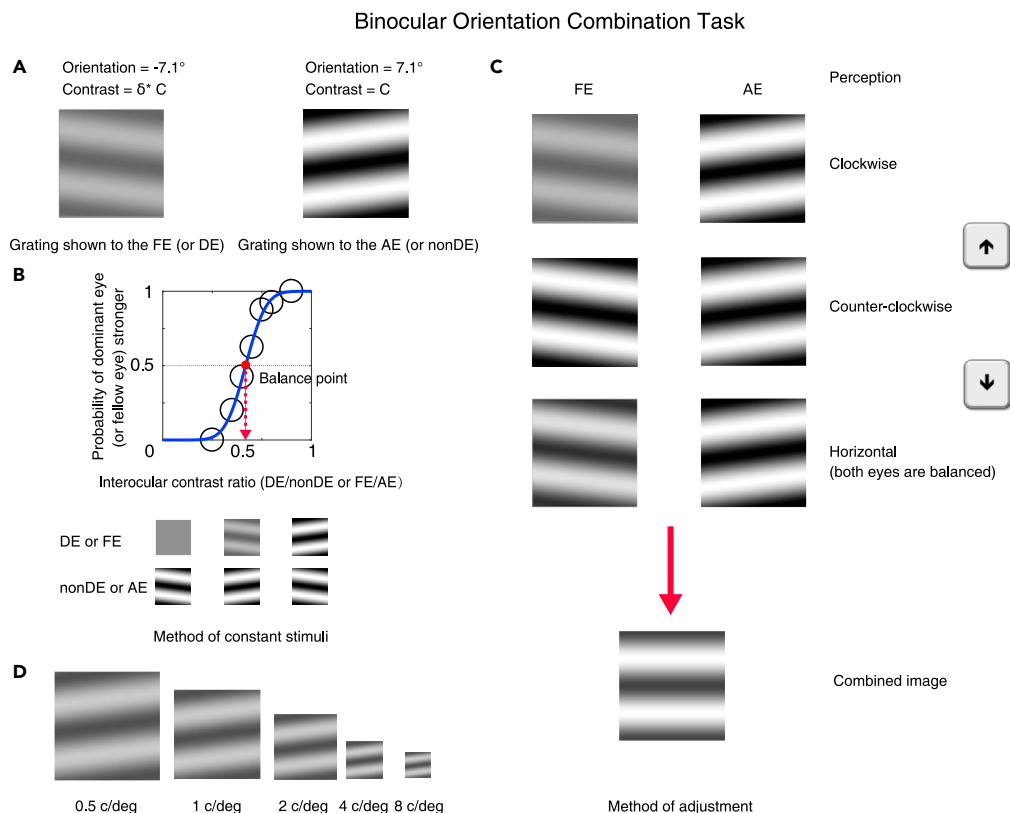
Recently, Wang et al. introduced a binocular combination test that overcomes the issue of spatial resolution and, therefore, is capable of assessing interocular balance for stimuli at higher spatial frequencies (Wang et al., 2019). The new test is a binocular orientation combination task. The original phase combination task, which is constrained by spatial resolution, displays sinusoidal gratings with no tilt. However, in the binocular orientation task, the gratings are slightly tilted. Wang et al. (2019) introduced this task because human adults could combine two slightly oriented gratings when the orientation difference was less than 20° (Spiegel et al., 2016; Yehezkel et al., 2016; Wang et al., 2019). The tilt of the gratings enabled the new test to overcome the issue of spatial resolution. They demonstrated that the task is accurate at mid to high spatial frequencies (up to 8 c/deg) in normal observers. Recently, Mao et al. used the orientation task to test adult amblyopes (Mao et al., 2020) and confirmed the findings of previous studies (Ding et al., 2013a, 2013b; Kwon et al., 2015) that binocular imbalance peaks at a higher spatial frequency (4 c/deg).

In the studies of Wang et al. and Mao et al., the authors implemented the method of the constant stimuli to display the stimuli (Mao et al., 2020; Wang et al., 2019), which in turn were shown at a fixed set of interocular contrast ratios. There were 280 trials (i.e., 15–20 min of testing time) for a measurement per spatial frequency. The interocular contrast ratio (i.e., balance point), where both eyes would contribute equally to binocular combination, was then estimated by fitting a psychometric function. This approach is essential in the laboratory where time would not be of concern. However, clinicians have limited time for each patient. To implement the orientation task in the clinic, we modified the orientation task by adopting the method of adjustment to make the test duration much shorter. In this task, subjects were required to adjust the level of the contrast of one eye's stimulus until the orientation of stimuli from both eyes was combined in a balanced way. In this new task, we had no predefined set of contrast for displaying the stimuli as we used in our laboratory measuring using the method of constant stimuli. We reduced the testing duration to 2–5 min per measurement for each spatial frequency. In this study, we examined how well the clinically convenient adjustment method captured binocular balance in normal and amblyopic observers relative to the laboratory-based method of constant stimuli. Moreover, we evaluated the test-retest reliability and replicability of the adjustment task. We conclude that it can reliably measure binocular imbalance in amblyopes across multiple spatial frequencies in merely about 7 min and provide clinicians with a much-needed tool to assess binocular imbalance of their patients quickly and efficiently (Figure 1).

## RESULTS

### Does the adjustment method replicate the previous findings of binocular imbalance in amblyopia?

Previous studies show that binocular imbalance gets worse as a function of spatial frequency (Chen et al., 2021; Kwon et al., 2015; Mao et al., 2020) in amblyopia. However, binocular balance has been found to be stable across spatial frequency in normal observers (Mao et al., 2020). In our study, we were able to replicate the findings of the previous studies in both normal and amblyopic participants. For each observer, we fitted a linear regression of  $|\log BP|$  as a function of spatial frequency (see Figures 2 and 3). This revealed the slope and y-intercept for each psychophysical test. For the normal observers, the slopes were relatively flat (i.e., near zero) for all tasks (Figure 2). This indicates that binocular balance is stable across spatial frequency in the normal observers. However, for the amblyopic observers, the slopes were all significantly higher than zero (one-sample t test,  $p < 0.05$ ; see Figure 3). This suggests that binocular imbalance increases as a function of spatial frequency.



**Figure 1. Stimuli and design**

FE = fellow eye, DE = dominant eye, AE = amblyopic eye, nonDE = non-dominant eye.

(A) The stimuli were two horizontal sinusoidal gratings with a tilt of  $\pm 7.1^\circ$ . They were shown separately to each eye. The grating shown to the dominant or the fellow eye had a lower contrast than that shown to the non-dominant or the amblyopic eye.

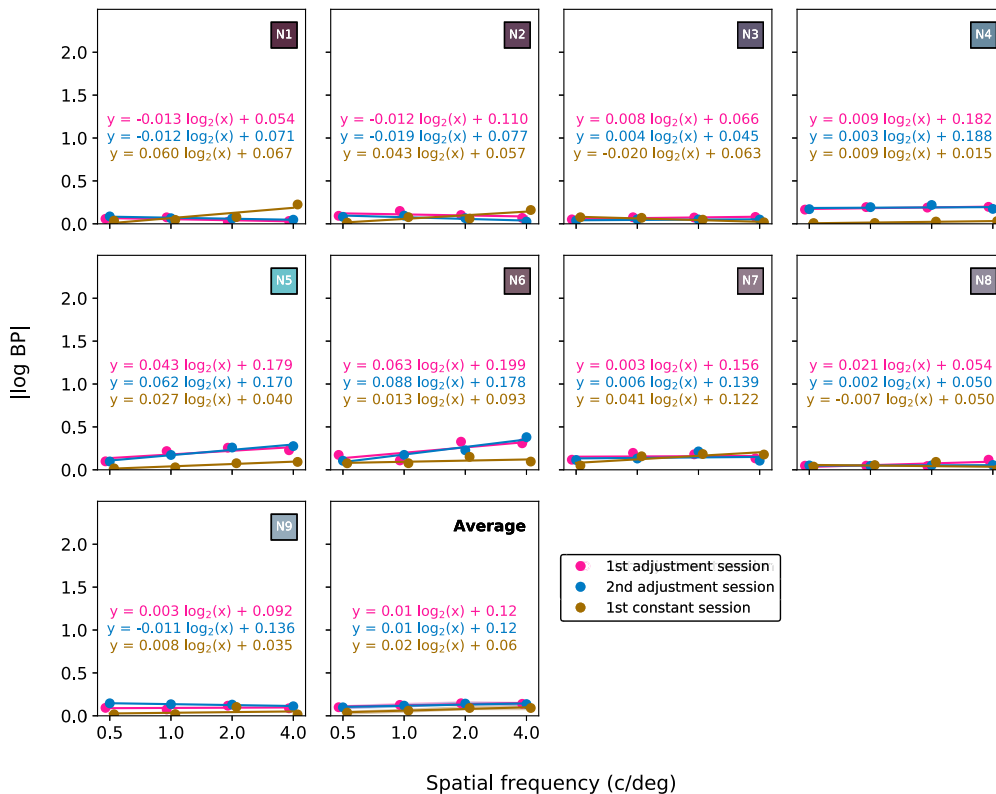
(B) A psychometric function (i.e., a cumulative Gaussian function of distribution) of a representative subject with normal vision. The y-axis indicates the probability in which the stimuli shown to the dominant eye (or fellow eye) is perceived to be stronger. The balance point (BP) is the contrast ratio of the two eyes that results in the perfect balance between the two eyes (50% probability of the dominant eye to be stronger). In the constant stimuli version, gratings shown to both eyes had pre-established contrasts (i.e., seven interocular contrast ratios). These are indicated by the unfilled circles in the psychometric function.

(C) For normal observers, the contrast of the gratings for the dominant eye was ranged from 0%–100%, whereas the contrast for the non-dominant eye was fixed at 50%. For amblyopes, the contrast of the grating for the fellow eye was ranged from 0% to 100%, and the contrast for the amblyopic eye was fixed at 100%. The interocular contrast ratio was ranged from 0 to 1 for the amblyopic observers, and 0 to 2 for the normal observers.

(D) Gratings shown in the method of adjustment. Using the keyboard, subjects were able to adjust the contrast of the grating shown to the fellow (or dominant eye). The contrast of the grating shown to the nondominant eye was maintained at 50% (normal observers), and the contrast of the amblyopic eye was kept at 100% (amblyopic observers; shown in the panel). Hence, in the method of adjustment (the novel version of the task), the subjects were able to modulate the contrast level of the grating shown to their dominant/fellow eye. In other words, the contrast of the grating shown to the dominant/fellow eye was not pre-established.

As an index of binocular imbalance, area under a curve (AUC) for each subject was computed. AUC was calculated between  $|\log \text{BP}|$  (log10 units in y-axis) and linear units of spatial frequency (c/deg in x-axis). The larger the area, the greater the binocular imbalance. Figure 4 shows how we computed AUC for two representative subjects (a normal observer, N3; and an amblyopic observer, A12). For a normal observer (subject N3), AUCs were close to zero at all three tasks. However, for an amblyopic observer (subject A12), AUCs were all above 3. We compared the AUCs across the three experimental sessions using one-way ANOVA (within-subject factor: experimental session). There was no significant difference for normal observers ( $F(2,12) = 2.17, p = 0.14$ ) and amblyopes ( $F(2,36) = 0.86, p = 0.43$ ). However,  $p > 0.05$  would not indicate that the null hypothesis is true. It would merely indicate that the means of each group do not differ by

Linear regression of balance points from 9 normal observers



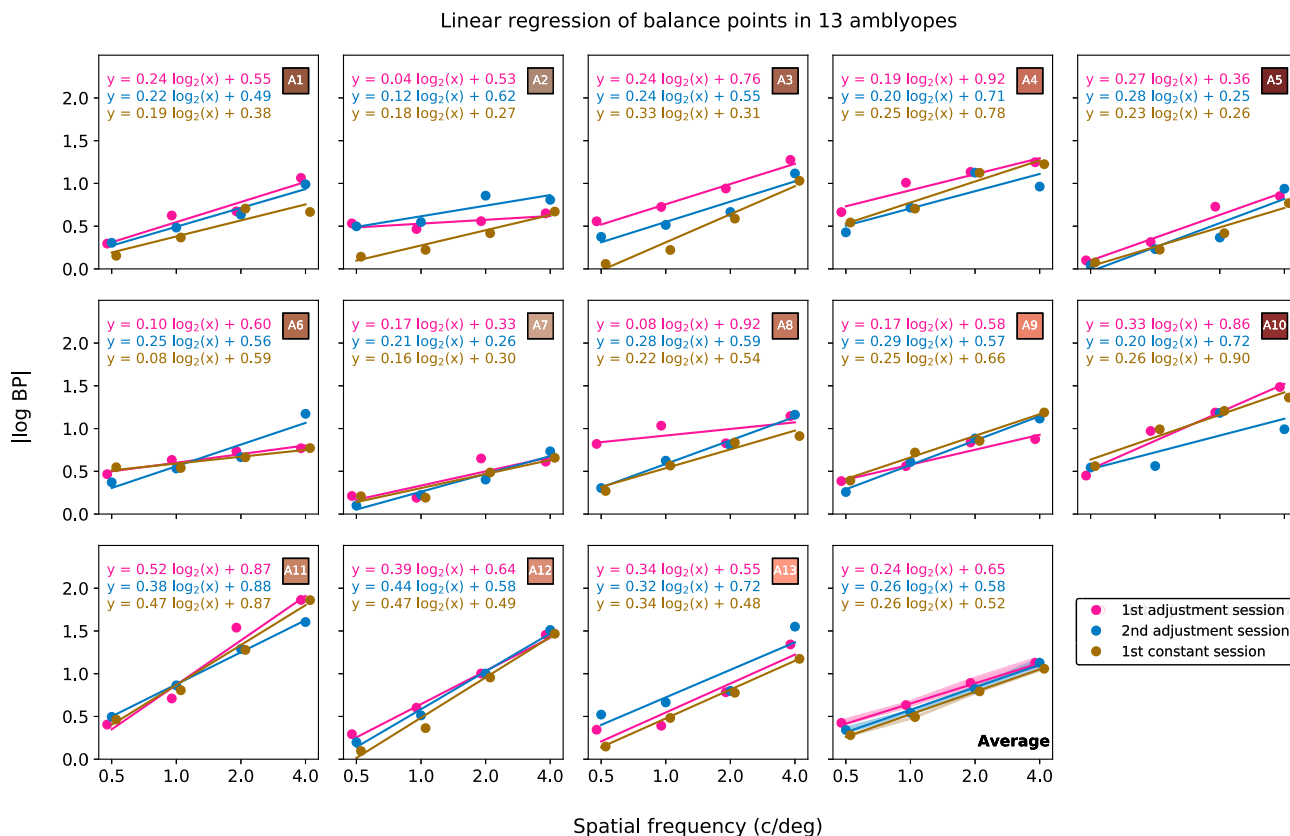
**Figure 2. Linear regression of  $|\log \text{BP}|$  as a function of spatial frequency in normal observers**

Cyan plots represent  $|\log \text{BP}|$  from the first session of the adjustment task. Blue plots represent  $|\log \text{BP}|$  from the second session of the adjustment task. Mustard plots represent  $|\log \text{BP}|$  from the one session of the constant stimuli method task (i.e., standard version). Subject labels for normal observers are shown in the top right corner of each panel.

more than two standard deviations. So, in the sections below, we will directly examine the test-retest reliability and replicability between the two methods, and between the two sessions of the adjustment method.

### Do data from the adjustment method agree to those from the constant stimuli method?

First, we investigated whether the data from the two methods were correlated. To do so, a correlation coefficient was computed via the Pearson test between AUCs from the one session of constant stimuli method and averaged AUCs across two adjustment sessions. The correlation was weak ( $r = 0.11$ ,  $p = 0.774$ ) for normal observers but robust ( $r = 0.88$ ,  $p < 0.001$ ) for amblyopic observers (see Figures 5A and 5C). However, a strong correlation does not guarantee a good replicability. To evaluate replicability, we analyzed the data with a Bland-Altman plot (see Figures 5B and 5D). For normal observers, the upper and lower limits of agreement (i.e., 95% limits of agreement) were  $-0.608$  and  $0.268$ . These limits represent the range within which the difference is most likely to fall for most observers. The mean difference between the two methods was  $-0.170$ . The AUC difference between the methods was found not to be significantly different from the 0 (one-sample t test:  $t(8) = 1.89$ ,  $p = 0.10$ ; see Figure 5B). For amblyopes, the upper and lower limits of agreement were  $-0.949$  and  $0.458$ , respectively. The mean difference between the methods was  $-0.245$ . The AUC difference between the methods was statistically significant from 0 (one-sample t test:  $t(12) = 2.46$ ,  $p = 0.030$ ). This significance is illustrated in Figure 5D, which shows that 0 in the y-axis does not overlap with 95% confidence interval of the AUC difference (predicted from t-distribution) and thereby suggests a statistical significance. This finding suggests because the AUC of adjustment is higher than that of the constant stimuli method by  $0.245$ , this mean difference might need to be adjusted when both methods are employed in the same setting. Nevertheless, the 95% limits of agreement for both normal and amblyopic observers are narrow enough in the clinical context. As shown in Figure 5, amblyopes can have a higher AUC than normal observers by around 3.



**Figure 3. Linear regression of  $|\log \text{BP}|$  as a function of spatial frequency in amblyopic observers**

Cyan plots represent  $|\log \text{BP}|$  from the first session of the adjustment task. Blue plots represent  $|\log \text{BP}|$  from the second session of the adjustment task. Mustard plots represent  $|\log \text{BP}|$  from the one session of the constant stimuli method task (i.e., standard version). Subject labels for amblyopes are shown in the top right corner of each panel.

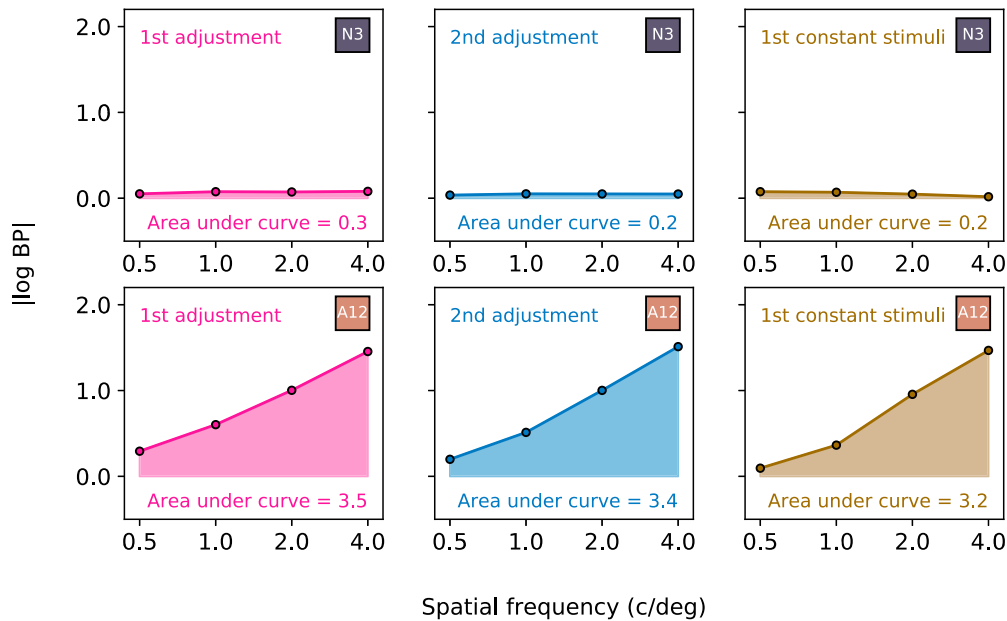
### How is the test-retest reliability and replicability between the two sessions of the adjustment method?

To begin with, we examined whether the data from the two sessions of the adjustment version were correlated. To do so, a correlation coefficient was computed via the Pearson test between AUCs from the two adjustment sessions. The correlation was robust for normal ( $r = 0.88$ ,  $p < 0.001$ ) and for amblyopic observers ( $r = 0.95$ ,  $p < 0.001$ ; Figures 6A and 6C). However, a good correlation would not ensure a strong replicability. To address replicability, we analyzed the data with a Bland-Altman plot (see Figures 6B and 6D). For normal observers, the upper and lower limits of agreement (i.e., 95% limits of agreement) were  $-0.147$  and  $0.118$ , respectively. These limits represent the range within which the difference is most likely to fall for most observers. The mean difference between the two methods was  $-0.015$ . It was not statistically different from the value of 0 (one-sample t-test:  $t(8) = 0.64$ ,  $p = 0.54$ ). For amblyopes, the upper and lower limits of agreement were  $-0.967$  and  $0.567$ , respectively. The mean difference between the methods was  $-0.200$ . The difference between the two methods was not found to be statistically significant from the value of 0 ( $t(12) = 1.84$ ,  $p = 0.091$ ). The 95% limits of agreement for both normal and amblyopic observers seem to be narrow in the clinical context. As shown in Figure 4, amblyopes can have a higher AUC than normal observers by around 3.

### Can the total testing time of the adjustment method be reduced more?

On average, each amblyopes spent 2 to 5 min per test at each spatial frequency. Because we tested four spatial frequencies (0.5, 1, 2, and 4 c/deg), we spent up to 20 min to measure binocular imbalance (i.e., AUC) for each patient. Would it be possible to reduce the testing duration by removing some conditions without compromising the reliability of the task?

Integrated binocular imbalance in representative subjects



**Figure 4. Area under a curve (AUC) as an index for binocular imbalance**

Cyan plots represent  $|\log BP|$  from the first session of the adjustment task. Blue plots represent  $|\log BP|$  from the second session of the adjustment task. Mustard plots represent  $|\log BP|$  from the one session of the constant stimuli method task (i.e., standard version).

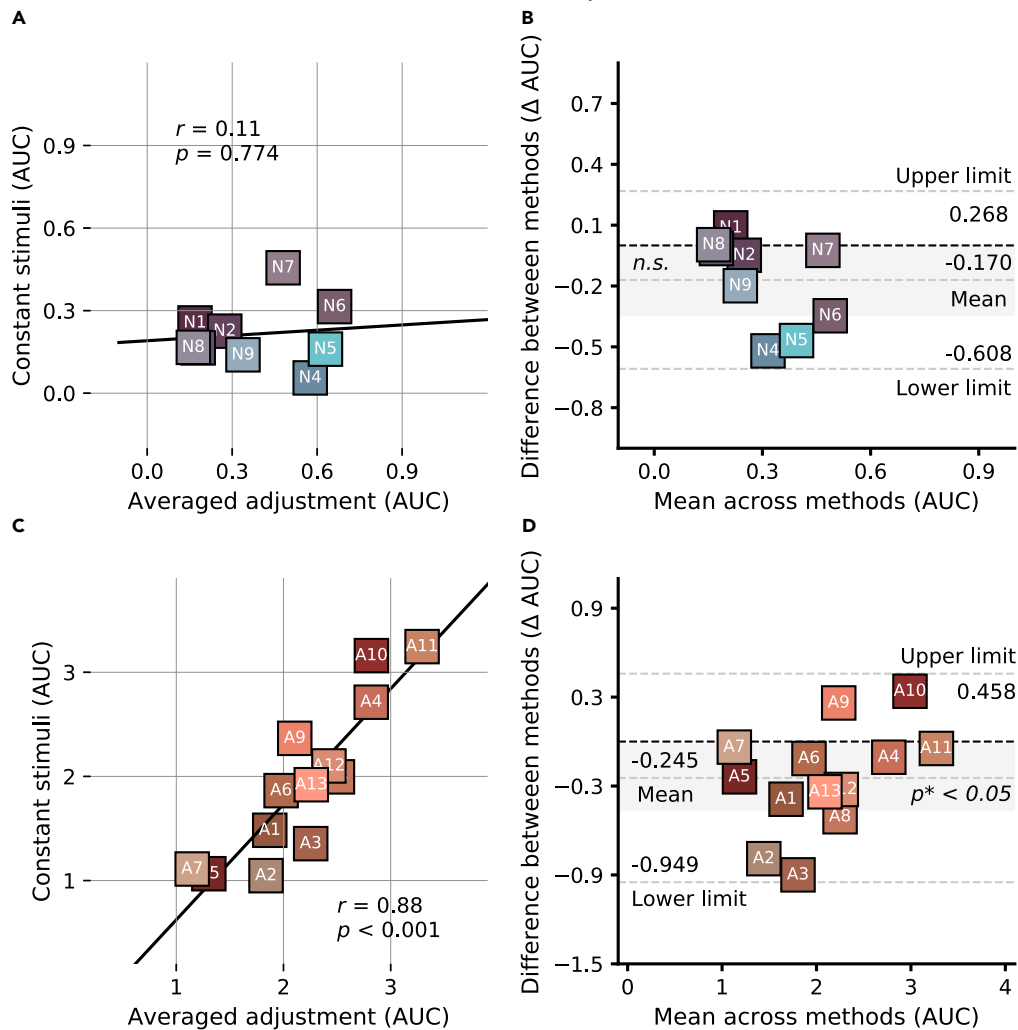
AUC depends strongly on the outer values of x-axis. In our design, the lowest and highest spatial frequencies that were measured were 0.5 and 4 c/deg. Therefore, we checked whether spatial frequencies such as 1 and 2 c/deg would be necessary in both normal adults and amblyopes. In normal observers, for both session of the adjustment test, a paired t test revealed that there was no significant difference ( $p$ 's > 0.14, Cohen's  $d$ 's < 0.75) between the AUCs from all SFs and those from the outer SFs (0.5 and 4 c/deg). In addition, the correlation between the AUCs from all SFs and those from the two outer SFs were highly correlated ( $r$ 's  $\geq 0.95$ ,  $p$ 's < 0.001; see Figures 7A and 7C). In amblyopes, for both sessions of the adjustment test, a paired t test revealed that there was no significant difference ( $p$ 's > 0.64, Cohen's  $d$ 's < 0.19) between the AUCs from all SFs and those from the outer SFs (0.5 and 4 c/deg). Moreover, the correlation between the AUCs from all SFs and those from the two outer SFs were highly correlated ( $r$ 's  $\geq 0.87$ ,  $p$ 's < 0.001; see Figures 7B and 7D).

Moreover, the test-retest reliability and repeatability of the adjustment test using AUCs computed from only 0.5 and 4 c/deg were remarkably similar with AUCs from all SFs (see Figure 8 for their correlations) in amblyopes. To see if the data were well correlated, we calculated the correlation coefficient, which was found to be significant ( $r = 0.75$ ,  $p = 0.003$ ; see Figure 8A). We further analyzed the data to examine the replicability of data by plotting a Bland-Altman plot. The upper and lower limits of agreement (i.e., 95% limits of agreement) were  $-0.979$  and  $0.732$ , respectively. The mean difference was  $-0.123$ . In addition, the difference was not found to be significantly different from zero (one-sample t-test,  $p = 0.090$ ; see Figure 8B, labeled as *n.s.*). These results indicate that the adjustment method is reliable when only at 0.5 and 4 c/deg was included in computing binocular imbalance. In short, this task would be able to reliably measure binocular imbalance in just 4–10 min (i.e., two test blocks) for amblyopes across spatial frequency.

## DISCUSSION

A recent study raises the question of whether solely measuring visual acuity to assess amblyopia is truly informative for clinical diagnosis (Chen et al., 2021). It shows that individuals whose amblyopia had been treated via a monocular patching therapy with supposedly a success had remaining deficits in their abilities to binocularly combine visual input, especially at high spatial frequencies. Physicians who prescribe a

Constant stimuli vs. adjustment methods

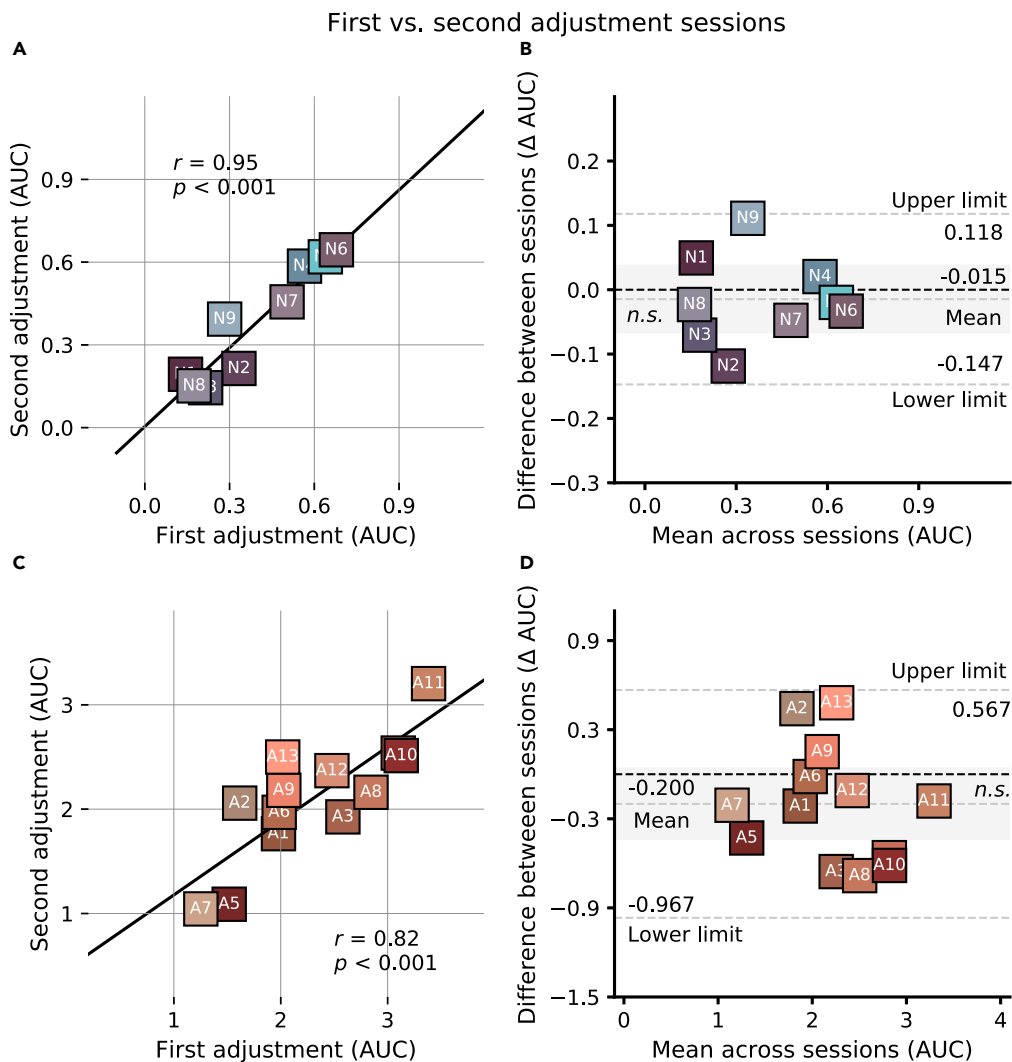


**Figure 5. Test-retest reliability and replicability between the constant stimuli and adjustment methods**

AUC from one session of the constant method and averaged AUCs across two adjustment sessions were analyzed for a comparison. Normal observers are labeled in cool colors (blue, navy etc.), whereas amblyopes are labeled in warm colors (red, brown, etc.).

- (A) Test-retest reliability computed by a Pearson's correlation test of the normal observers ( $n = 9$ ). The solid black line indicates the best-fit linear regression from the correlation.
- (B) Bland-Altman plot of the normal observers. The x-axis represents mean AUCs across the two methods for each subject. The y-axis represents the difference of AUCs between the two methods for each subject. The middle-dashed silver line represents mean difference of AUC. If the mean difference is significantly different from 0 (i.e., one-sample t-test), the agreement between the methods can be deemed as poor. In this panel, a one-sample t-test indicates that the mean difference is not significantly different from 0 ( $y = 0$  is shown as a black dashed line), which is labeled as *n.s.* (not significant) in the panel. The gray-colored area, which represents the 95% confidence interval of the AUC difference between methods predicted from a t-distribution, overlaps with the black dashed line ( $y = 0$ ). 95% limits of agreement are shown by the upper and lower silver dashed lines; they represent the range within which the difference is most likely to fall for most observers. The wider the limits of agreement, the larger the measurement variability between the tasks.
- (C) Test-retest reliability computed by a Pearson's correlation test of the amblyopic observers ( $n = 13$ ).
- (D) Bland-Altman plot of the amblyopic observers. The AUC difference between the methods is significantly different from 0 (labeled as  $p^* < 0.05$ ); this is shown by the fact that the black dashed line ( $y = 0$ ) does not overlap with 95% confidence interval of the AUC difference (gray-shaded area).





**Figure 6. Test-retest reliability and replicability between the two sessions of the adjustment method**

The four panels are illustrated similarly to those in Figure 5.

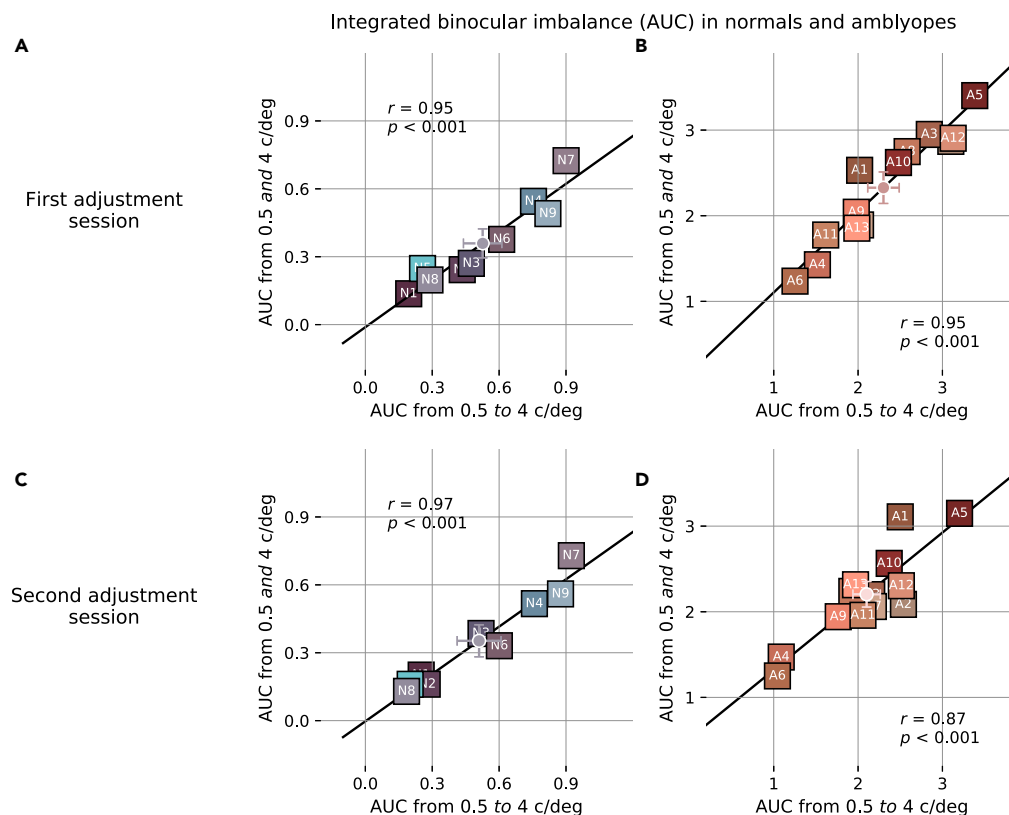
(A) Test-retest reliability computed by a Pearson's correlation test of the normal observers ( $n = 13$ ). The solid black line indicates the best-fit linear regression from the correlation.

(B) Bland-Altman plot of the normal observers. It is illustrated similarly to those in Figure 5. The AUC difference between the methods is not significantly different from 0 (labeled as *n.s.* = not significant); the black dashed line ( $y = 0$ ) does overlap with 95% confidence interval of the AUC difference (grey-shaded area).

(C) Test-retest reliability computed by a Pearson's correlation test of the amblyopic observers ( $n = 13$ ).

(D) Bland-Altman plot of the amblyopic observers. The AUC difference between the methods is not significantly different from 0 (labeled as *n.s.* = not significant); the black dashed line ( $y = 0$ ) does overlap with 95% confidence interval of the AUC difference (gray-shaded area).

standard patching therapy monitor the visual acuity of the amblyopic eye as the main index for the severity of amblyopia. In other words, even if there was a minimal difference in visual acuity of the eyes in these treated individuals, the binocular deficit still remained, which is typically observed in untreated amblyopic patients (Ding et al., 2013a, 2013b; Kwon et al., 2015; Mao et al., 2020). It has also been shown to be correlated with deficits in stereopsis from amblyopia (Han et al., 2018; He et al., 2018; Li et al., 2011). The fact that the supposedly treated individuals still have residual deficits in their binocular balance at higher spatial frequencies illustrates that the current means to assess the severity of amblyopia, which relies on visual acuity difference between the eyes, is clearly lacking, and that treatment which targets binocular improvement should be recommended along with a typical monocular patching therapy to ensure a more thorough

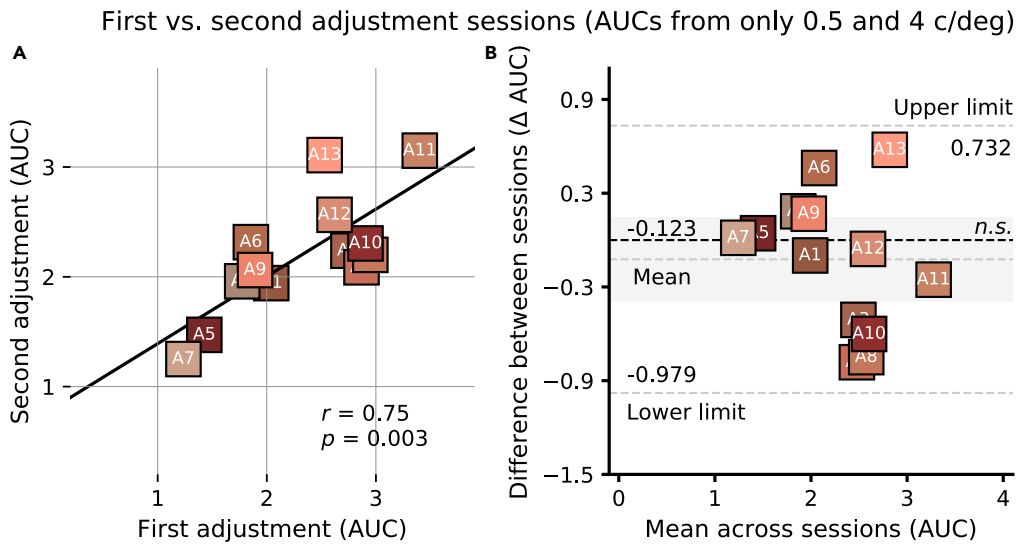


**Figure 7. AUC calculated based on four frequencies (0.5–4 c/deg, x-axis) vs. two frequencies (0.5 and 4 c/deg, y-axis). In all panels, the average and the standard errors (error bars) are shown. In addition, the solid line indicates the best-fit linear regression for each correlation.**

- (A) First adjustment session of normal observers.
- (B) First adjustment session of amblyopes.
- (C) Second adjustment session of normal observers.
- (D) Second adjustment session of amblyopes.

visual recovery. In short, both monocular visual acuity and binocular balance should be essential criteria of assessment for amblyopia.

However, there has been no test that could measure binocular balance at multiple spatial frequencies quickly enough to be implemented in the clinic. Wang et al. introduced the binocular orientation task that uses the laboratory-based method of constant stimuli (Wang et al., 2019); this task has been used widely in recent studies (Chen et al., 2021; Mao et al., 2020). In this test, a predefined set of contrasts is established before data collection (Wang et al., 2019). However, it can take about 15–20 min for a measurement of binocular balance at each spatial frequency (280 trials each), thereby amounting to more than 30 min total to compute an areal measure (AUC) of binocular balance as a function of spatial frequency. Another method that has been used is a dichoptic chart that displays band-pass filtered letters (Kwon et al., 2015). In this test, at least 60 points are accumulated to measure the balance at each spatial frequency, amounting to 6–7 min of testing for each spatial frequency after subjects have completed practice sessions (Kwon et al., 2015). However, because this test presents different letters to each eye dichoptically, it mainly assesses binocular rivalry rather than binocular combination, which is a process by which both eyes fuse to form a functional binocular image and significantly affects everyday life (Johansson et al., 2014; Suttle et al., 2011). The testing time of our newly introduced clinical orientation task is even more brief and as precise. It can measure binocular balance of a patient at multiple spatial frequencies in about 7 min. In our study, subjects were asked to adjust the level of the contrast of one eye’s stimulus until the stimuli from both eyes were balanced (i.e., method of adjustment). There were 16 trials for each testing block (for each spatial frequency). As an index of binocular imbalance, we computed the area under a curve (AUC) of  $|\log BP|$  as a



**Figure 8. Test-retest reliability and replicability of AUCs from amblyopes**

They were computed from just two spatial frequencies (0.5 and 4 c/deg). The two panels are illustrated similarly to those in Figure 5.

(A) Test-retest reliability computed by a Pearson's correlation test ( $n = 13$ ).

(B) Bland-Altman plot. The AUC difference between the methods is not significantly different from 0; the black dashed line ( $y = 0$ ) does overlap with 95% confidence interval of the AUC difference (gray-shaded area).

function of spatial frequency. Our data of the areal index from the new test successfully captured the relationship between binocular imbalance and high spatial frequency that typically characterizes amblyopia.

Using the AUC, we evaluated the precision of the adjustment test relative to its standard counterpart (i.e., constant stimuli). Although the range of AUC difference between methods in amblyopes was not wide, there was a fixed bias of +0.245 between the AUCs of the adjustment method and those of the constant stimuli method. Should a clinician decide to use both methods in a single setting or study, one might need to subtract 0.245 AUC from the adjustment method or add 0.245 AUC to the constant stimuli method for a proper standardization (see Figure 5D).

The AUC difference between the adjustment method and the constant stimuli method in amblyopes might not seem large if the reader compares this discrepancy with the AUC difference obtained from normal observers ( $-0.245$  vs.  $-0.170$  AUC). However, our analysis indicates that the AUC difference between the methods in amblyopes is statistically significant from 0 (one-sample t-test,  $p = 0.030$ ). We suspect that the between-methods AUC difference could be attributed two causes. The first cause could be the choice of the stimuli's starting contrast of the dominant/fellow eye (for normal and amblyopes, respectively) in the adjustment method. During the measurement, the contrast of the grating shown to the nondominant eye was maintained at 50% (normal observers), and the contrast of the amblyopic eye was kept at 100% (amblyopic observers). The starting contrast was either at 100% or 1% for the dominant/fellow eye. The subjects were asked to modulate the contrast of the stimuli by either 5% (fine) or 10% (coarse regulation). Because the two eyes of normal observers would usually be equally balanced, they would reach a balanced contrast ratio when the dominant eye was shown with 50% contrast. For this reason, the number of ascending and descending trials would roughly be the same. However, the amblyopes' vision would be much more imbalanced. Even if the amblyopic eye would be shown with contrast at 100%, the fellow eye would need a contrast much less than 50% to reach an appropriate balance point (ex. 20% contrast for the dominant eye). This would lead to the amblyopic subject to perform fewer ascending trials and more descending trials. One solution to resolve this issue would be to establish an appropriate starting contrast of the fellow eye for each amblyopic observer during a training session so that the number of ascending and descending trials would be equal during the actual experiment. The second cause for the AUC difference could be due to the constant error stemming from the observer's tendency to repeatedly be conservative or liberal when adjusting the contrast of the grating to reach a balance during the adjustment method. The constant error stems from the starting contrast level of the stimulus (Jones 1989). It is possible that the two causes could have interacted and amplified the margin of the error, thereby contributing to widening the AUC difference

between the adjustment and constant stimuli methods. This error could perhaps be resolved with a training session that includes a sample test block that incorporates correct/incorrect responses and a more appropriate starting contrast of the stimulus.

Next, we evaluated the test-retest reliability and replicability of the adjustment method itself (see [Figure 6](#)). To do so, we tested all subjects twice using the adjustment method. The correlation between the two sessions was excellent in both normal and amblyopic observers. In addition, the test-retest difference in AUC from the adjustment test was not wide (less than 1). Our findings indicate that the adjustment method is highly reliable and replicable across multiple testing sessions.

Clinicians can measure binocular imbalance at all four spatial frequencies (0.5, 1, 2, and 4 c/deg) using the adjustment version if their time is unlimited for each of their patient. In this case, it can take about 8–20 min to complete the test for each patient. However, a testing duration that exceeds 10 min can be too long in the clinic. Because areal measures could heavily be depended on the outer x-axis values (i.e., 0.5 and 4 c/deg), we performed an additional analysis using AUCs computed only from 0.5 to 4 c/deg and determined whether the data from 1 to 2 c/deg might be superfluous. We found that the AUCs from the outer spatial frequencies were similar to AUCs computed from all four spatial frequencies (see [Figure 7](#)). This indicates that the testing duration can be reduced without compromising the quality of the data. In addition, the correlation and test-retest difference were excellent and quite similar to those from AUCs computed across all four spatial frequencies (see [Figure 8](#)). Therefore, our additional analysis indicates that the orientation task using the method of adjustment can quantify the extent of binocular imbalance from low (e.g., 0.5 c/deg) to mid (e.g., 4 c/deg) frequencies in just 4 to 10 min if the researcher decides to test only the two outer spatial frequencies (i.e., 0.5 and 4 c/deg) rather than all four spatial frequencies (0.5, 1, 2, and 4 c/deg).

Quantitative assessment can greatly benefit in designing a unique treatment regimen for each patient ([Black et al., 2011](#); [Li et al., 2013](#); [Narasimhan et al., 2012](#)). Our orientation task, which is a quantitative measure, can provide a more adequate assessment than those from the readily available, more qualitative tests such as Worth 4-dot test, the Bagolini lenses, and OXO tests. These tests only offer a binary categorization of whether patients have suppression or not. Moreover, they do not evaluate the relationship between binocular imbalance and spatial frequency. Furthermore, studies show that patients find these tests challenging ([Piano and Newsham 2015](#)). The task difficulty can reduce the test-retest reliability and replicability. In this study, we show that a psychophysical task, namely binocular orientation combination via the method of adjustment, can be used to measure binocular balance of patients across spatial frequency quickly and reliably. Our method is easy and intuitive. More importantly, it replicates the findings of previous studies on binocular imbalance in amblyopia ([Ding et al., 2013a, 2013b](#); [Kwon et al., 2015](#); [Mao et al., 2020](#)). In this study, only adults had been tested. However, we believe that a younger population can perform this task easily because the task has a low cognitive demand. In the future, one could perhaps refine the adjustment task to make it more engaging and help the children maintain their focus.

### Limitations of the study

This study has a few limitations. First, the sample size (9 controls and 13 amblyopes) is not very large. Second, we were not able to collect data at higher spatial frequencies for amblyopes because the patients were not able to perform the task reliably at 8 c/deg or above. Third, most of the amblyopic participants were anisometropic (A3-A11). So, our results regarding the task's reliability might need to be verified in patients who have different types of amblyopia or other binocular disorders.

### 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 SUBJECT DETAILS](#)
  - Participants
- [METHOD DETAILS](#)

- Experimental design
- Stimuli
- Procedure
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Balance point (BP)
  - Data analysis
- **ADDITIONAL RESOURCES**

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.103652>.

## ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China Grant (NSFC31970975), the Natural Science Foundation for Distinguished Young Scholars of Zhejiang Province, China (LR22H120001), and the National Key Research and Development Program of China Grant (2020YFC2003800) to JZ, the Zhejiang Medical Health Science and Technology Project (2020KY656) to ZH and the Canadian Institutes of Health Research Grant (CCI-125686) to RH. The sponsor or funding organizations had no role in the design or conduct of this research.

## AUTHOR CONTRIBUTIONS

S.M., Y.M., R.H., and J.Z. conceived the experiments. Y.M. and S.C. performed the experiments. S.M., Y.M., S.C., Z.H., R.H., and J.Z. analyzed and interpreted the data, and wrote the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## DECLARATION OF INTERESTS

There is no conflict of interest.

Received: September 23, 2021

Revised: October 25, 2021

Accepted: December 13, 2021

Published: January 21, 2022

## REFERENCES

- Adrian, J., Le Brun, J., Miller, N.R., Sahel, J.-A., Saillant, G., and Bodaghi, B. (2019). Implications of monocular vision for racing drivers. *PLoS One* 14, e0226308.
- Beylerian, M., Hess, R.F., Matonti, F., Denis, D., Chavane, F., and Reynaud, A. (2020). Interocular suppressive interactions in amblyopia depend on spatial frequency. *Vis. Res.* 168, 18–28.
- Black, J.M., Thompson, B., Maehara, G., and Hess, R.F. (2011). A compact clinical instrument for quantifying suppression. *Optom. Vis. Sci.* 88, E334–E343.
- Brainard, D.H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436.
- Chen, S., Min, S.H., Cheng, S.H., Xiong, Y., Yu, X., Wei, L., Mao, Y., Hess, R.F., and Zhou, J. (2021). Binocular visual deficits at mid to high spatial frequency in treated amblyopes. *iScience* 24, 102727.
- Dane, A., and Dane, S. (2004). Correlations among handedness, eyedness, monocular shifts from binocular focal point, and nonverbal intelligence in university mathematics students. *Perceptual Mot Skills* 99, 519–524.
- Ding, J., Klein, S.A., and Levi, D.M. (2013a). Binocular combination in abnormal binocular vision. *J. Vis.* 13, 14.
- Ding, J., Klein, S.A., and Levi, D.M. (2013b). Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model. *J. Vis.* 13, 13.
- Ding, J., and Sperling, G. (2006). A gain-control theory of binocular combination. *Proc. Natl. Acad. Sci. U S A* 103, 1141–1146.
- Han, C., He, Z.J., and Ooi, T.L. (2018). On sensory eye dominance revealed by binocular integrative and binocular competitive stimuli. *Invest. Ophthalmol. Vis. Sci.* 59, 5140–5148.
- He, Z.J., Ooi, T.L., and Su, Y.R. (2018). Perceptual mechanisms underlying amodal surface integration of 3-D stereoscopic stimuli. *Vis. Res.* 143, 66–81.
- Heinen, T., and Vinken, P.M. (2011). Monocular and binocular vision in the performance of a complex skill. *J. Sports Sci. Med.* 10, 520.
- Hess, R.F., Thompson, B., and Baker, D.H. (2014). Binocular vision in amblyopia: structure, suppression and plasticity. *Ophthalmic Physiol. Opt.* 34, 146–162.
- Holopigian, K., Blake, R., and Greenwald, M.J. (1986). Selective losses in binocular vision in anisometropic amblyopes. *Vis. Res.* 26, 621–630.
- Huang, C.-B., Zhou, J., Lu, Z.-L., Feng, L., and Zhou, Y. (2009). Binocular combination in anisometropic amblyopia. *J. Vis.* 9, 17.
- Huang, C.-B., Zhou, J., Zhou, Y., and Lu, Z.-L. (2010). Contrast and phase combination in binocular vision. *PLoS One* 5, e15075.
- Johansson, J., Pansell, T., Ygge, J., and Seimyr, G.Ö. (2014). Monocular and binocular reading performance in subjects with normal binocular vision. *Clin. Exp. Optom.* 97, 341–348.
- Jones, L.A. (1989). Matching forces: constant errors and differential thresholds. *Perception* 18, 681–687.
- Kwon, M., Wiecek, E., Dakin, S., and Bex, P. (2015). Spatial-frequency dependent binocular imbalance in amblyopia. *Sci. Rep.* 5, 17181.

- Levi, D.M., Harwerth, R.S., and Smith, E.L. (1980). Binocular interactions in normal and anomalous binocular vision. *Doc.Ophthalmol.* 49, 303–324.
- Li, J., Thompson, B., Deng, D., Chan, L.Y.L., Yu, M., and Hess, R.F. (2013). Dichoptic training enables the adult amblyopic brain to learn. *Curr. Biol.* 23, R308–R309.
- Li, J., Thompson, B., Lam, C.S.Y., Deng, D., Chan, L.Y.L., Maehara, G., Woo, G.C., Yu, M., and Hess, R.F. (2011). The role of suppression in amblyopia. *Invest. Ophthalmol. Vis. Sci.* 52, 4169–4176.
- Mansouri, B., Thompson, B., and Hess, R.F. (2008). Measurement of suprathreshold binocular interactions in amblyopia. *Vis. Res.* 48, 2775–2784.
- Mao, Y., Min, S.H., Chen, S., Gong, L., Chen, H., Hess, R.F., and Zhou, J. (2020). Binocular imbalance in amblyopia depends on spatial frequency in binocular combination. *Invest. Ophthalmol. Vis. Sci.* 61, 7.
- Narasimhan, S., Harrison, E.R., and Giaschi, D.E. (2012). Quantitative measurement of interocular suppression in children with amblyopia. *Vis. Res.* 66, 1–10.
- Piano, M., and Newsham, D. (2015). A pilot study examining density of suppression measurement in strabismus. *Strabismus* 23, 14–21.
- Reynaud, A., and Hess, R.F. (2016). Is suppression just normal dichoptic masking? suprathreshold considerations. *Invest. Ophthalmol. Vis. Sci.* 57, 5107–5115.
- Spiegel, D.P., Baldwin, A.S., and Hess, R.F. (2016). The relationship between fusion, suppression, and diplopia in normal and amblyopic vision. *Invest.Ophthalmol. Vis.Sci.* 57, 5810–5817.
- Suttle, C.M., Melmoth, D.R., Finlay, A.L., Sloper, J.J., and Grant, S. (2011). Eye-hand coordination skills in children with and without amblyopia. *Invest. Ophthalmol. Vis. Sci.* 52, 1851–1864.
- Wang, Y., He, Z., Liang, Y., Chen, Y., Gong, L., Mao, Y., Chen, X., Yao, Z., Spiegel, D., and Qu, J. (2019). The binocular balance at high spatial frequencies as revealed by the binocular orientation combination task. *Front. Hum. Neurosci.* 13, 106.
- Yehezkel, O., Ding, J., Sterkin, A., Polat, U., and Levi, D. (2016). Binocular combination of stimulus orientation. *R. Soc. Open Sci.* 3, 160534.
- Zhou, J., Huang, P.C., and Hess, R.F. (2013). Interocular suppression in amblyopia for global orientation processing. *J. Vis.* 13, 19.
- Zhou, J., Liu, R., Feng, L., Zhou, Y., and Hess, R.F. (2016). Deficient binocular combination of second-order stimuli in amblyopia. *Invest. Ophthalmol. Vis. Sci.* 57, 1635–1642.
- Zhou, J., Liu, R., Zhou, Y., and Hess, R.F. (2014). Binocular combination of second-order stimuli. *PLoS One* 9, e84632.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Experimental models: Organisms/strains</i>		
9 normal observers (age: $23.6 \pm 2.24$ years, mean $\pm$ SD; four males)	Recruited at Wenzhou Medical University	N/A
13 amblyopes (age: $26.8 \pm 6.7$ years; eight males; see <a href="#">Table S1</a> for details)	Recruited at Wenzhou Medical University	N/A
<i>Software and algorithms</i>		
MatlabR2016bv9.1.0	MathWorks	<a href="https://www.mathworks.com/products/matlab.html">https://www.mathworks.com/products/matlab.html</a>
Psychtoolbox extension v3.0.14	<a href="#">Brainard (1997)</a>	<a href="http://psychtoolbox.org/">http://psychtoolbox.org/</a>
RStudio	RStudio	<a href="https://www.rstudio.com/">https://www.rstudio.com/</a>
<i>Other</i>		
MacBook Pro 2017	Apple, Inc.	<a href="https://www.apple.com/mac/">https://www.apple.com/mac/</a>
ASUS monitor (PG279Q)	AsusTek Computer Inc.	<a href="https://www.asus.com/Displays-Desktops/Monitors/All-series/">https://www.asus.com/Displays-Desktops/Monitors/All-series/</a>
GOOVIS (AMOLED display)	NED Optics	<a href="https://goovis.en.alibaba.com/company_profile.html">https://goovis.en.alibaba.com/company_profile.html</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jiawei Zhou ([zhoujw@mail.eye.ac.cn](mailto:zhoujw@mail.eye.ac.cn)).

#### Materials availability

No new unique reagents were developed by this study.

#### Data and code availability

All data reported in this study will be available from the lead contact author (Jiawei Zhou: [zhoujw@mail.eye.ac.cn](mailto:zhoujw@mail.eye.ac.cn)) upon request.

The paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact (Jiawei Zhou: [zhoujw@mail.eye.ac.cn](mailto:zhoujw@mail.eye.ac.cn)) upon request.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Participants

We tested 9 normal observers (age:  $23.6 \pm 2.24$  years, mean  $\pm$  SD; four males) and 13 amblyopic adults (age:  $26.8 \pm 6.7$  years; eight males; see [Table S1](#) for details). Observers wore their normal optical correction if required. A written informed consent was obtained from each of them before the start of the test. This study was in line with the Declaration of Helsinki and was approved by the International Review Boards at Wenzhou Medical University. All subjects consented to the study and completed each experimental condition twice, each of which was separated by at least 24 h.

**Apparatus.** On a 13-inch MacBook Pro (Apple, Inc., Cupertino, CA, USA), experiments were conducted using MATLABR2016b (v9.1.0MathWorks, Inc., Natick, MA, USA) with the PsychToolBox extension ([Brainard, 1997](#); [Pelli, 1997](#)). The stimuli were shown through gamma-corrected head-mounted goggles

(GOOVIS, AMOLED display, NED Optics, Shenzhen, China) with a refresh rate of 60 Hz, a resolution of 2560 × 1600 and a maximal luminance of 150 cd/m<sup>2</sup>. They were displayed dichoptically; one eye was shown with a slightly different stimulus than the other eye.

## METHOD DETAILS

### Experimental design

In this study, we assessed whether the data obtained from an orientation task using the clinically convenient adjustment method were consistent with those obtained from the laboratory-based method of constant stimuli, and therefore, determined whether the adjustment method was reliable. Thus, subjects were asked to complete three sessions of the experiment: 1) one session of the standard version (i.e., constant stimuli method; Figures 1A and 1B), 2) first session of the adjustment version (Figure 1D) and 3) second session of the adjustment version. By doing so, we were able to examine whether binocular imbalance from the more abbreviated, clinical orientation task (i.e., adjustment version) was similar to that of the laboratory standard task. Also, we were able to evaluate test-retest reliability (i.e., Pearson's correlation test) and replicability (Bland-Altman Plot) of the proposed adjustment version. Each experiment block at each spatial frequency takes about 15 min to complete in the constant stimuli method. The adjustment method, which is what we propose in this paper to be used clinically, takes about 2–5 min per experimental block at each spatial frequency (0.5, 1, 2 and 4 c/deg).

### Stimuli

There were two displayed sinusoidal gratings in two different orientations. In the first configuration, the left eye was shown a grating with an orientation of +7.1° counter-clockwise relative to the horizontal axis, whereas the right eye was shown one with an orientation of –7.1° counter-clockwise position. In the second configuration, the left eye was shown a grating with an orientation of –7.1° counter-clockwise relative to the horizontal axis, whereas the right eye was shown one with an orientation of +7.1° counter-clockwise position. We averaged the binocular imbalance measures from these two configurations to remove potential positional bias.

*Method of constant stimuli (standard but time-consuming, 280 trials, 15 min per spatial frequency).* Seven contrast ratios between the eyes were selected for each subject based on the pilot data. For normal observers, the contrast of the gratings for the dominant eye ranged from 0–100%, whereas the contrast for the non-dominant eye was fixed at 50%. For amblyopes, the contrast of the grating for the fellow eye ranged from 0 to 100%, and the contrast for the amblyopic eye was fixed at 100%. The eye dominance of the normal observers was established by a hole-in-the-hand test (Dane and Dane, 2004).

Each orientation combination configuration and contrast ratio were tested for 20 times. There were 280 trials total (2 orientation configurations × 7 contrast ratios × 20 repetitions) involved in the testing of one spatial frequency. The order of the configuration and contrast ratios were randomized. The subjects perceived a fused grating from the combination of the two monocular gratings of equal but opposite orientation. Subjects reported via keyboard whether they perceived a clockwise or counterclockwise fused grating.

*Adjustment method (new, quick and apt for clinic, 16 trials, 2–5 min per spatial frequency).* For amblyopes, we fixed the contrast of the tilted grating presented to the amblyopic eye at 100%. When testing normal subjects, we fixed the contrast of the grating presented to the nondominant eye at 50%. The contrast of the grating presented to the fellow eye in amblyopes or that to the dominant eye in normal was adjusted by observers until the fused grating between the eyes appeared to be horizontal. The starting contrast was set as 1% in half of the trials and 100% in the other half. Each orientation configuration was repeated for 4 times, so there were 16 trials total (2 orientation configurations × 2 starting contrasts × 4 repetitions) in one test session for each spatial frequency. The mean interocular contrast ratio (fellow eye/amblyopic eye or dominant eye/nondominant eye) was calculated based on these 16 trials in testing one spatial frequency. Two alternatives for relative contrast step size of the grating were provided to each observer: 10% for coarse regulation and 5% for fine regulation. However, most subjects chose 10% for regulation.



### Procedure

In both the constant stimuli and adjustment versions of the orientation task, there were two parts in each trial. The first part was the alignment phase, during which we ensured the subjects achieved a proper fusion of the stimuli's that were dichoptically displayed. In each screen (one for each eye), a segment of a cross appeared. Two segments formed a cross. However, they were not fully aligned to form a proper cross. Therefore, subjects were asked to align them with the keyboard. After the alignment phase, a blank screen followed for 500 ms, then the test phase began. The positions of the aligned stimuli were recorded and used in the test phase. In which, a horizontal sinusoidal grating with a different orientation (titled either clockwise or counterclockwise relative to the horizontal axis) appeared in each screen. Subjects reported with the keyboard.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Balance point (BP)

For the constant stimuli method, we plotted the probability of the subjects perceiving the percept shown to the dominant eye (or fellow eye) as a function of contrast ratios between the eyes (i.e., interocular contrast ratios, dominant/nondominant or fellow/amblyopic eyes). Next, we fitted these points with a cumulative Gaussian distribution function to estimate the value of interocular contrast ratio that results in the probability of 50%, where both eyes contribute equally to binocular combination. We refer to this interocular contrast ratio as the balance point (BP). BP was transformed into the absolute values of  $\log_{10}$  units (i.e.,  $|\log BP|$ ) to indicate the extent of binocular imbalance. Therefore,  $|\log BP|$  of 0 indicates a balanced binocular vision. The higher the  $|\log BP|$ , the more imbalanced the binocular vision.

### Data analysis

We used Python and R for data analysis. We computed area under a curve (AUC) as an index for binocular imbalance; this is the integral of  $|\log BP|$  as a function of spatial frequency (0.5 to 4 c/deg). AUC was calculated between  $|\log BP|$  plots ( $\log_{10}$  units in y-axis) and linear units of spatial frequency (c/deg in x-axis). With the AUCs, we performed Pearson's correlation test to measure test-retest reliability. A strong correlation indicates that the data across the two methods are related in a similar fashion across subjects. However, it does not mean good replicability. To assess replicability, we used Bland-Altman plots. These plots enabled us to directly measure the test-retest difference. The lower the test-retest difference, the more robust the replicability.

## ADDITIONAL RESOURCES

There is no additional forum that provides additional information about this study, which is not part of a clinical trial.