Vision Research 247 (2026) 108873

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Impact of crowding on visual appearance and performance in amblyopia☆

Ângela Gomes Tomaz[a,*], Wolf Harmening[b], Dennis M. Levi[a], Bilge Sayim[c]

[a] Herbert Wertheim School of Optometry & Vision Science, University of California Berkeley, Minor Hall, 380 Minor Ln, Berkeley, CA 94720, United States of America
[b] Department of Ophthalmology, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany
[c] Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP), Département d'Études Cognitives (DEC), École normale supérieure (ENS), CNRS, PSL University, 29 rue d'Ulm, 75005 Paris, France

A R T I C L E I N F O

Keywords:
Shape
Spatial vision
Appearance
Crowding
Amblyopia
Similarity

A B S T R A C T

Individuals with amblyopia have been reported to perceive stimuli as distorted with the amblyopic eye, and to have stronger foveal crowding than controls without amblyopia. Previous studies on visual appearance in amblyopia often lacked a control group and did not adjust stimulus size to account for visual acuity deficits. Here, we investigated whether observers with amblyopia perceived stimuli as more distorted and whether foveal crowding influenced visual appearance more strongly than in controls, while controlling for visual acuity differences. Isolated and flanked target letters were presented foveally to the dominant/fellow and non-dominant/amblyopic eye. Stimuli were scaled relative to each eye's acuity threshold and presented at 1.0× and 1.5× threshold size. Observers replicated target appearance by selecting squares on a novel grid interface. Results showed that accuracy, structural similarity (SSIM) and perceived similarity (PSS) were lower for observers with amblyopia compared to controls (Group), for small compared to large stimuli (Size), and for flanked compared to isolated targets (Crowding). Interestingly, a significant interaction between Group and Crowding was observed only for accuracy and SSIM, but not for PSS. These results showed group differences despite controlling for acuity, suggesting that factors beyond visual acuity, such as stronger foveal crowding, contribute to visual appearance differences in amblyopia compared to controls.

1. Introduction

Amblyopia is a neurodevelopmental disorder of the visual system caused by a lack of correlated visual input, typically during the sensitive period of development from birth to about 6–8 years of age (Von Noorden, 1981; Worth, 1903). Amblyopia is usually classified according to the underlying factors presumed to have caused the uncorrelated visual input. Amblyogenic factors include anisometropia (i.e., significant unequal refractive power between the two eyes), strabismus (i.e., misalignment of the eyes), form deprivation (i.e., obstruction of vision such as due to cataracts) or both strabismus and anisometropia (Gabai & Zeppieri, 2023). Clinically, it is characterized by an interocular difference in visual acuity (VA; the ability to resolve spatial detail) of two or more chart lines (i.e., at least 0.2 logMAR difference in interocular acuity) despite best optical correction (Wright, 2006) and no observable ocular pathology. Amblyopia is one of the main causes of vision loss in both children and adults (Birch, 2013; Gunnlaugsdottir et al., 2008). Persons with amblyopia display a vast array of low level deficits (e.g., contrast sensitivity, Levi & Harwerth, 1980; Zele et al., 2007; position acuity, Levi, 2021; stereopsis, see Levi et al., 2015 for a review on stereopsis in amblyopia) and higher level deficits (e.g., counting, Sharma et al., 2000; decision-making, Farzin & Norcia, 2011; visual search, Tsirlin et al., 2018; attention, Popple & Levi, 2008; see also Verghese et al., 2019 for a review on attentional deficits in amblyopia and Levi, 2020 for a general review), due to the significant physiological alterations in their visual pathways (e.g., Kiorpes, 2019; Kiorpes & Daw, 2018; Kiorpes & McKee, 1999; Wiesel & Hubel, 1963, 1965). One such case where observers with amblyopia differ from controls is visual crowding (e.g., Flom et al., 1963; Hariharan et al., 2005; Hess & Jacobs, 1979; Stuart & Burian, 1962).

In visual crowding target identification is impaired by neighboring stimuli (i.e., flankers) (Bouma, 1970; Flom et al., 1963; Herzog & Sayim, 2022; Herzog et al., 2016; Levi, 2008; Parkes et al., 2001; Whitney & Levi, 2011). Crowding limits perception across the entire visual field,

☆ This article is part of a Special issue entitled: 'Amblyopia' published in Vision Research.

* Corresponding author at: 486-488 Minor Hall, University of California, Berkeley, Berkeley, CA 94720, United States of America.
E-mail addresses: angela.gomestomaz@gmail.com (Â.G. Tomaz), wolf.harmening@ukbonn.de (W. Harmening), dlevi@berkeley.edu (D.M. Levi), bilge.sayim@cnrs.fr (B. Sayim).

https://doi.org/10.1016/j.visres.2026.108873
Received 1 May 2025; Received in revised form 10 June 2026; Accepted 26 June 2026

though it differs between peripheral and foveal regions. In the normal fovea, the crowding extent (i.e., the critical target-flanker distance) is proportional to target size such that if the stimulus size is scaled, performance remains the same (Levi, Klein, et al., 2002). In this case, foveal crowding is indistinguishable from contrast masking (Levi, Klein, et al., 2002). However, near the foveal resolution limit, with smaller target-flanker spacings, crowding appears to not scale with target size and is more easily distinguished from contrast masking (Danilova and Bondarko, 2007). Furthermore, when optical factors are controlled for (e.g., by employing optical aberration correction with an adaptive optics scanning laser ophthalmoscope, AOSLO), the critical spacing for foveal crowding is much smaller (≈0.75–1.3 arcminutes of target-flanker edge-to-edge spacing) than previously measured (Coates et al., 2018). Many effects of foveal crowding (e.g., Malania et al., 2007; Sayim et al., 2008, 2010) have also been found in the periphery (Manassi et al., 2012, 2013).

In amblyopia, stronger foveal crowding has been found for observers with strabismic and mixed amblyopia compared to observers with anisometropic amblyopia and controls (Bonneh et al., 2004, 2007; Hess & Jacobs, 1979; Polat et al., 2005; Song et al., 2014; Waugh & Fronius, 2024; Wong et al., 2005). Further, greater foveal crowding extent has been found in the amblyopic eye of observers with amblyopia compared to controls (Flom et al., 1963; Hariharan et al., 2005; Hess & Jacobs, 1979; Levi, Hariharan, et al., 2002; Stuart & Burian, 1962), with significant crowding differences between the amblyopic vs. the fellow eye in children (with strabismic and mixed amblyopia, Greenwood et al., 2012; Kalpadakis-Smith et al., 2022), as well as adults (Levi, Hariharan, et al., 2002; Stuart & Burian, 1962). Although earlier studies suggested that crowding in amblyopia is not more extensive than in controls when visual acuity deficits are accounted for (Flom et al., 1963; Simmers et al., 1999), subsequent research indicated that crowding effects in amblyopia do not seem to depend on the eccentric fixation (i.e., when the point of fixation is outside the fovea, Schor & Hallmark, 1978) and the visual acuity deficit of the observer (Bonneh et al., 2004; Hess et al., 2001; Hess & Jacobs, 1979; Levi, Hariharan, et al., 2002; Levi & Klein, 1985; Song et al., 2014). As in the normal periphery, crowding extent in strabismic amblyopia does not depend on target size (Hariharan et al., 2005; Levi & Carney, 2011; Levi, Hariharan, et al., 2002; Simmers et al., 1999), though in anisometropic amblyopia (as in the normal fovea), the critical distance has been found to be proportional to target size (Levi & Carney, 2011). In previous studies, foveal crowding in amblyopia has been found to be dependent on stimulus features such as spatial frequency (Polat et al., 2005), and target/flanker contrast polarity (Hess et al., 2001). However, it does not seem to be orientation specific as in the normal fovea (Hariharan et al., 2005).

Crowding is typically investigated with performance-based tasks in which observers are asked to discriminate a specific feature of the target (e.g., its orientation) or to report its identity (e.g., when the target is a letter). In performance-based tasks, observers often respond by selecting from a number of allowed response alternatives (e.g., letters) which are then scored binarily as correct or incorrect. Although fast, such quantification of performance can be restrictive as observers are constrained to only a few (often just two) response categories, and cannot report any percept that does not fit the available categories. Recently, a number of appearance-based methods have been employed to enable responses beyond the typical categories when measuring performance. For example, appearance-based research on crowding in observers without visual or cortical deficits has shown that a crowded target is not only more difficult to identify, but also strongly differs in its appearance compared to uncrowded targets (Coates et al., 2017; Sayim & Wagemans, 2017). In one study, Sayim and Wagemans (2017) asked observers to draw isolated and crowded letters, numbers and letter-like symbols presented at 0°, 6° and 12° eccentricity gaze-contingent with unlimited stimulus presentation times using eye tracking. Drawings were analyzed for omission, addition, truncations, extensions, translations, rotations of elements and shape distortions relative to the target. Using these fine-grained error categories, previously unreported crowding errors were identified and quantified, revealing new details of crowded target appearance and how appearance changes as a function of crowding strength.

In amblyopia, appearance under crowding has also been studied using drawings. Pugh (1958) asked observers to identify, describe and draw letters (T, F, E, H, X, Z, A, N, O, C, D, U) as they were perceived, decreasing letter size with correct identification (from the biggest size on a clinical visual acuity chart to the smallest letter size the observer could read). The letters were shown isolated, as part of a single chart row, and as part of a full visual acuity chart. Observers' reports revealed distortions of letter appearance such as perceiving the letters blurred, rounded, having indefinite edges, and with thinner strokes than the presented stimuli. Asymmetrical distortions in letter shape were common in observers with strabismus (e.g., an observer with a left amblyopic eye and a small degree of esotropia perceived the right side of the letters sharp, while the left side appeared 'jagged' or 'smudged'). Interestingly, there were large differences in target appearance between observers (see also Molaei et al., 2025), but within each observer, the distortions were consistent across different letters.

Distortions in amblyopia have also been found in the appearance of circle stimuli (Sireteanu et al., 1993) and gratings (Barrett et al., 2003; Bradley & Freeman, 1985; Hess et al., 1978), although not under crowding. In these studies, as in Pugh's (1958) study, there were qualitative differences in perception across observers. For example, in Hess et al.'s (1978) study, some observers perceived the grating bars thinner, while others perceived them more 'jumbled'. Importantly, despite differences in spatial distortions in the two eyes, some observers' contrast sensitivity (i.e., a performance-based measure) was similar in the two eyes. Further, Barrett et al. (2003) instructed observers to sketch perceived distortions of gratings when viewing them with the amblyopic eye, using only the fellow eye for their responses. While most observers showed significant spatial distortions (67%), a substantial proportion (33% of observers) did not, and these distortions did not seem to depend on the severity or etiology of amblyopia. These findings highlight the importance of investigating appearance alongside performance to further the understanding of amblyopia.

Here, we investigated the appearance of isolated and crowded stimuli in amblyopia using a novel method to capture appearance. Specifically, we asked whether observers with amblyopia would show differences from controls for isolated or crowded letters when visual acuity differences were accounted for. Visual acuity differences were accounted for by adjusting target sizes to individual thresholds. Observers replicated the appearance of isolated and flanked letters as seen by the dominant/fellow (DE/FE) and non-dominant/amblyopic eye (NDE/AE) at and above visual acuity threshold sizes. Near threshold sizes were chosen to ensure foveal crowding and not contrast masking for both groups.

2. Methods

2.1. Participants

Adult observers with unilateral amblyopia were recruited from the UC Berkeley Optometry Eye Clinic and screened by an optometrist prior to inclusion in the study. The screening included auto-refraction, as well as subjective refraction to measure observers' refractive error and find the optimal correction to achieve best corrected visual acuity. Observers with amblyopia met the following inclusion criteria: a clear amblyogenic factor (e.g., strabismus, if any tropia was present; anisometropia, i.e., interocular difference in the spherical-equivalent of at least 1.0 diopter; or mixed, if both), as well as an interocular visual acuity difference with best-corrected vision of at least 0.2 logMAR in the absence of ocular pathology or history of eye surgery (except to correct strabismus or remove cataracts). The only exception to the logMAR acuity criterion was AMBFOV_6 as her interocular acuity difference at the time of screening was 0.14 logMAR (only 3 letters short of our criterion of 0.2 logMAR difference, i.e., less than 1 chart line) and her age made her a good option to best age-match the control and amblyopia groups. Adult control observers were recruited with flyers around the UC Berkeley campus. Control observers with normal or corrected-to-normal visual acuity participated in the study. To avoid the impact of uncorrected low order optical aberrations in the findings, only observers whose VA was 0 logMAR (20/20) or better in each eye were included in the final sample.

Forty-six observers were tested (Table S1 in the *Supplementary Materials* shows the reason for exclusion of participants who initially performed the experiment, but were removed from the final sample). The final sample comprised twenty-nine naïve observers: fifteen adult controls (6 males, age range: 18–33, mean age: 22 ± 4 years) and fourteen adults with unilateral amblyopia (5 males, age range: 19–59, mean age: 30 ± 14 years) (6 with anisometropia, 5 with strabismus, 2 mixed, and 1 form deprivational) (clinical information of each observer with amblyopia is provided in Table A).

The current study complied with the ethical standards of the

Table A
Patient's clinical information.

Participant code	Age/Sex	Amblyopia type/eye	Refraction (at a distance)	Snellen acuity	Anisometropia amplitude	Cover test	Comment
AMBFOV_3	31/M	Anisometropia/OS	OD: −4.00 -0.50 × 111 OS: Plano	OD: 20/16 OS: 20/80 + 2	4.25 D	Dist.: 2 Δ EP Near: 4 Δ EP	
AMBFOV_6	19/F	Anisometropia/OD	OD: Plano OS: −1.50 -0.75 × 170	OD: 20/32–1 OS:20/25 + 1	3.75 D	Dist.: 4 Δ LEP Near: 8 Δ LEP	
AMBFOV_12	59/M	Anisometropia/OS	OD: +1.00–1.50 × 085 OS: +5.75–1.50 × 077	OD: 20/20 + 2 OS: 20/40 + 1	4.75 D	Dist.: 2 Δ XP Near: 4 Δ XP	
AMBFOV_15	26/M	Anisometropia/OS	OD: +3.00–4.50 × 006 OS: +3.50–3.25 × 172	OD: 20/20 OS: 20/63	1.13 D	Dist.: Ortho Near: 2 Δ XP	
AMBFOV_17	57/F	Strabismus/OD	OD: Plano −0.50 × 155 OS: −0.25 -0.50 × 015	OD: 20/160 OS: 20/20	0.25 D	Dist.: 8 Δ RET Near: 12 Δ RET	Small right hypertropia for both distance and near cover test, used her personal, more recent, glasses (see Rx)
AMBFOV_20	19/M	Form deprivation/OS	OD: −9.25 -2.00 × 180 OS: −7.25 -2.75 × 175	OD: 20/32 OS: 20/63 + 2	1.63 D	Dist.: 18 Δ LXT Near: 25 Δ LXT	Surgery for cataract removal in the left eye around 7–8 years old
AMBFOV_23	32/M	Strabismus/OS	OD: Plano −0.50 × 180 OS: −0.50 -1.00 × 180	OD: 20/20 OS: 20/63–2	0.75 D	Dist.: Ortho Near: 8 Δ LET	
AMBFOV_29	19/F	Strabismus/OD	OD: +4.25 OS: +3.75	OD: 20/80 + 2 OS: 20/15–2	0.50 D	Dist.: 5 Δ RXT Near: 7 Δ RXT	Microstrabismus (from 4 Δ BO test); Eccentric fixation of ~1 degree in the TSLO)
AMBFOV_34	28/F	Mixed/OS	OD: Plano OS: +8.00	OD: 20/16–3 OS: 20/125–2	8.00 D	Dist.: 2 Δ LET Near: 2 Δ LET	Microstrabismus (from 4 Δ BO test)
AMBFOV_42	21/F	Mixed/OS	OD: +1.00 OS: +4.50	OD: 20/15–2 OS: 20/30	3.50 D	Dist.: Ortho Near: 2 Δ EP	Microstrabismus (from 4 Δ BO test)
AMBFOV_43	22/F	Anisometropia/OS	OD: −1.50 -0.5 × 110 OS: +3.00–3.00 × 180	OD: 20/20–3 OS: 20/40–1	3.25 D	Dist.: Ortho Near: Ortho	Microstrabismus (from 4 Δ BO test)
AMBFOV_44	46/F	Anisometropia/OD	OD: +2.50–0.75 × 110 OS: −1.00 - 0.75 × 047	OD: 20/50–2 OS: 20/20	3.50 D	Dist.: XP Near: XP	Microstrabismus (from 4 Δ BO test)
AMBFOV_45	22/F	Strabismus/OS	OD: Plano −1.00 × 175 OS: Plano −0.75 × 175	OD: 20/20 OS: 20/32–2	0.13 D	Dist.: 6 Δ EP Near: 6 Δ EP	Microstrabismus (from 4 Δ BO test)
AMBFOV_46	21/F	Strabismus/OS	OD: +5.00 OS: +5.00	OD: 20/25–2 OS: 20/40	0 D	Dist.: 4 Δ XP Near: 3 Δ XP	Microstrabismus (from 4 Δ BO test)

Note. M: Male; F: Female; OS - Left eye; OD: Right eye; Δ: Prism; Anisometropia amplitude was calculated as the absolute value of spherical equivalent difference between the OS and OD (and is given in diopters). Spherical equivalent was calculated as the spherical component plus half of the cylindrical component. In the cover test column, the number indicates the eye deviation amplitude in prism diopters (Δ) and the first letter (L or R) refers to the eye (left or right, respectively). XT or ET refer to exotropia or esotropia, respectively. XP or EP refer to exophoria and esophoria, respectively.

Declaration of Helsinki and was conducted with the approval (2010-09-2139) of the Institutional Review Board (IRB) of the University of California, Berkeley, where data collection took place. Written informed consent was obtained from all participants prior to the start of the experiment. Participants were monetarily compensated for their participation ($15 per hour).

2.2. Apparatus

A ViewSonic VX2250WM display with a pixel resolution of 1920 × 1080 pixels and 248.3 µm of pixel pitch was used for the first 5 participants. The monitor was a non-tactile and anti-glare thin film transistor (TFT) LCD with a 60 Hz refresh rate. An anti-glare TFT LCD DELL monitor (U2412Mb) with a resolution of 1920 × 1200 pixels and 269.2 µm of pixel pitch (refresh rate 60 Hz) was used for the remaining 24 participants. Both displays were gamma corrected in the experiment and used for the visual acuity assessment, stimulus display and data collection using Python 3.6.12 and PsychoPy 2020.2.10 (Peirce, 2007). A chinrest was used to ensure the proper viewing distances throughout the experiment.

2.3. Stimuli

Stimuli consisted of 5 letters (E, T, F, H, L). The letters were created on a 9 × 9 square grid used also as an interface to capture appearance (as in Gomes Tomaz et al., 2022). The five letters were chosen because they could be properly created and depicted on the grid (i.e., without round and/or diagonal elements, e.g., C, D, V, K). The letters were created on the central 25 squares (5 squares wide x 5 squares high) of the 9 × 9 square grid (Fig. 1A). The stimulus grid was not shown to observers during the stimulus presentation. Target letters were presented isolated (Fig. 1A, left) or flanked on all sides by black bars (bar width was one square on the grid and ⅕ of bar height) (Fig. 1A, right). Stimulus size was manipulated to 1.0 and 1.5 x the observer's minimum angle of resolution (MAR) (i.e., each square on the grid was the MAR or 1.5 x the MAR) (small and large size, respectively) (see *2.4. Procedure* section). The 9 × 9 response grid was always larger than the stimulus (the retinal sizes subtended by the response grid are shown in Table B). Stimuli were black (ViewSonic: 4.33 cd/m^2, Dell: 3.06 cd/m^2), presented in the center of the screen on a white (ViewSonic: 70.74 cd/m^2, Dell: 37.34 cd/m^2) background surrounded by a gray (ViewSonic: 33.73 cd/m^2, Dell: 22.54 cd/m^2) fixation ring.

2.4. Procedure

2.4.1. Visual acuity assessment

Prior to the start of the experiment, the observer's visual acuity (VA)

Table B
Retinal sizes in degrees subtended by the 9 × 9 response grid.

Eye tested	Control group			Amblyopia group		
	Min	*Max*	*M (SD)*	*Min*	*Max*	*M (SD)*
Dominant / Fellow eye	1.44	2.69	1.94 (0.35)	1.60	3.79	2.21 (0.62)
Non-dominant/ amblyopic eye	1.55	2.72	2.00 (0.29)	2.27	11.96	4.63 (2.37)

Note. *Min* - minimum, *Max* - maximum, *M* - mean, *SD* - standard deviation of the mean.

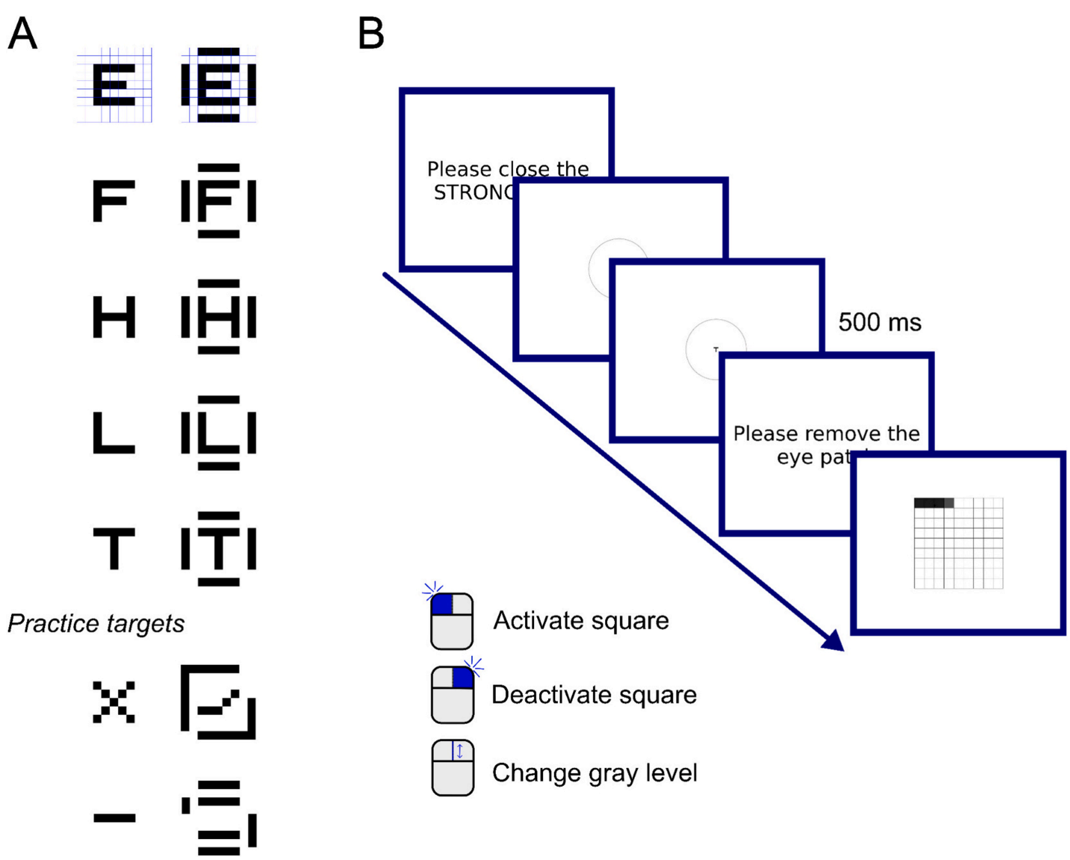

Fig. 1. A) Isolated (left) and flanked (right) targets used in the experiment and during practice. The grid behind the letter E is shown for illustration purposes. No grid was shown during the target presentation. B) Trial procedure. After occluding one of the eyes, observers fixated the center of the screen, followed by target presentation for 500 ms. Observers reported target appearance by selecting squares with mouse clicks and adjusting their gray levels with the mouse wheel. The images are not to scale.

and minimum angle of resolution (MAR) were measured using the Freiburg Visual Acuity and Contrast Test, FrACT (Bach, 1996; version 2021–12). Observers were placed 2.20 m away from the screen and their head was supported with a chinrest. Visual acuity was measured twice for each eye (while the other eye was patched) with 18 trials of isolated Sloan letters presented for 30 s in the center of the screen. To familiarize observers with the task, in particular observers with amblyopia who might see the letters more distorted, the dominant eye was always assessed first. Observers' task was to identify the letter as quickly and accurately as possible by pressing the keyboard key corresponding to the perceived letter. There was a 30 s time limit for responses (observers were not informed about the time limit). No feedback for correct/incorrect responses was given. The best visual acuity value recorded for each eye was used to set the viewing distance for each block in the experiment.

2.4.2. Main experiment: capturing letter appearance

In the main experiment, observers' task was to report the appearance of the stimuli. Observers started each trial by occluding either the dominant/fellow or non-dominant/amblyopic eye according to a prompt displayed on screen. After eye patching, a fixation ring was shown in the center of the screen. Observers were instructed to fixate its center and to press the spacebar when not blinking to initiate target presentation. Upon the spacebar press the isolated or flanked target letter either at 1.0 or 1.5 times the observer's MAR was presented for 500 ms. After target offset, a prompt to remove the eye occluder was presented. Observers removed the occluder and pressed the spacebar to be presented with the response grid (Gomes Tomaz et al., 2022). To capture target appearance, observers selected (black) and de-selected squares (white as the background) with mouse clicks (left and right mouse key, respectively) and adjusted their gray levels by scrolling the mouse wheel (Fig. 1B). There were 9gray levels (ViewSonic: 5.26 cd/m^2, 7.75 cd/m^2, 11.62 cd/m^2, 16.95 cd/m^2, 23.82 cd/m^2, 31.51 cd/m^2, 40.81 cd/m^2, 51.40 cd/m^2, 62.65 cd/m^2, Dell: 5.44 cd/m^2, 8.18 cd/m^2, 10.95 cd/m^2, 14.41 cd/m^2, 17.71 cd/m^2, 21.09 cd/m^2, 26.03 cd/m^2, 30.10 cd/m^2, 34.00 cd/m^2) available in between black (ViewSonic: 4.33 cd/m^2, DELL: 3.06 cd/m^2) and white (ViewSonic: 70.74 cd/m^2, DELL: 37.34 cd/m^2). Gray levels were chosen to approximate perceptual linearity from black to white. Over 97% of responses (97.46% flanked, 98.67% isolated) fell within the low-luminance range (3.06–16.95 cd/m^2), where screen luminance differed minimally between displays (mean difference 0.77 cd/m^2 for black and the four lowest-contrast gray levels). Reporting the same gray level for multiple squares was facilitated by having the mouse on top of a square with the desired gray level and pressing the 'C' keyboard key for 'copy'. After the gray level was copied, clicking with the left mouse key selected squares with that gray level. When finished using the same gray level, pressing the 'B' keyboard key for 'back' or 'black' would return the interface to the default setting (i.e., black squares selected by pressing the left mouse key). The response phase was binocular to ensure that all observers (especially observers with amblyopia) were able to use the response grid and report stimulus appearance as accurately as possible. After reporting the appearance of the stimulus, observers pressed the spacebar to move on to the next trial. The selected squares, their gray levels and the response time were recorded. No feedback was given on their responses.

Before the experiment, observers performed 2 practice trials per eye (one isolated and one flanked) (Fig. 1A, bottom). Observers always practiced with the dominant/fellow eye first to ensure that they understood the task and how to use the response grid. The practice targets were not used in the experiment.

The full experiment consisted of 2 blocks (one per eye), and their order was counterbalanced across participants. Each block comprised 20 trials where each target letter was presented once per crowding (isolated, flanked) and size (1.0 x MAR, 1.5 x MAR) condition. The trial order was randomized. Observers took breaks between blocks while the experimenter(s) adjusted the viewing distance by changing the position of the observers' desk relative to the screen. The adjustment of the viewing distance ensured that each square on the grid always subtended the 1.0 or 1.5 x the MAR on the retina of the individual observer's eye being tested. The mean viewing distances in each eye block per group are shown in Table C and the retinal sizes subtended by the stimuli in each eye and size condition are shown in Table D. The experiment ended when all trials were completed. The full experiment took approximately 1 h.

3. Analyses

Grid responses were assembled for visual inspection and analyzed regarding their accuracy (hits and correct rejections), structural similarity to the target stimulus (SSIM; Wang et al., 2004) and perceived response-target similarity (PSS). Statistical analyses were conducted in R 4.4.2. Separate generalized linear mixed effects models (glmm) were used to predict accuracy, SSIM, and PSS using the *glmmTMB* R package (version 1.1.11; Brooks et al., 2017; McGillycuddy et al., 2025). In all analyses, model assumptions were checked with the *DHARMa* package (version 0.4.7; Hartig, 2024), except for collinearity, which was checked with the *performance* package (version 0.12.4; Lüdecke et al., 2021). To confirm that the appropriate distribution was used, the expected distribution of the simulated scaled residuals was plotted against the observed distribution from the model, and a Kolmogorov-Smirnov test was performed. To check for linearity between predictors and the link-transformed response, the fitted values were plotted against the residuals. To further check for possible systematic misfits, a quantile regression on the residuals was performed against the predicted values. Residuals were also plotted across the levels of each categorical predictor, and Kolmogorov-Smirnov tests for uniformity and Levene tests for homogeneity of variance were conducted. As significant response variability might be expected between groups (control vs. amblyopia) and crowding conditions, dispersion was modelled accordingly to account for potential heteroscedasticity. To test whether there were more simulation outliers than expected, an outlier test based on an exact binomial test with approximate expectations was performed. A dispersion test was also conducted to test if the simulated dispersion was equal to the observed dispersion. Collinearity of the factors was evaluated by checking the variance inflation factor (VIF), and confirming that all values were smaller than 5. Zero inflation was also checked. Normality of the random effects was checked by plotting the theoretical quantiles against the sample quantiles, as well as with a Shapiro-Wilk test.

Model candidates included a null model containing only participants as a random effects factor, as well as models augmented with all independent variables (group, crowding condition, size, eye, eye order) as main fixed effects, and a model augmented with the group x crowding interaction. All models compared in each analysis had the same random effects and dispersion model formula. Models were compared to each other with likelihood ratio tests (LRT) to check whether they were significantly different. Models that met all assumptions and had the lowest Akaike Information Criterion (AIC; Matuschek et al., 2017) were preferred. As dispersion was modelled directly, standard r^2 measures do not provide reliable estimates and are therefore not reported. For ease of interpretation, model coefficients (estimates and standard errors) are

Table C
Experimental viewing distances (in mm) in each eye block and observer group.

Eye tested	Control			Amblyopia		
	Min	*Max*	*M (SD)*	*Min*	*Max*	*M (SD)*
Dominant / Fellow eye	2069	3867	2922 (552)	1470	3481	2650 (697)
Non-dominant/ amblyopic eye	2050	3307	2801 (361)	464	2456	1392 (538)

Note. *Min* - minimum, *Max* - maximum, *M* - mean, *SD* - standard deviation of the mean.

Table D
Retinal sizes (in degrees) subtended by the stimuli for both groups in each eye, crowding and size condition.

Crowding condition	Size (relative to resolution threshold)	Control group						Amblyopia group					
		Dominant eye			Non-dominant eye			Fellow eye			Amblyopic eye		
		Min	*Max*	*M (SD)*	*Min*	*Max*	*M (SD)*	*Min*	*Max*	*M (SD)*	*Min*	*Max*	*M (SD)*
Isolated	1.0	0.04	0.07	0.05 (0.01)	0.04	0.08	0.06 (0.01)	0.04	0.11	0.06 (0.02)	0.06	0.33	0.13 (0.07)
	1.5	0.06	0.11	0.08 (0.01)	0.06	0.11	0.08 (0.01)	0.07	0.16	0.09 (0.03)	0.09	0.50	0.19 (0.10)
Flanked	1.0	0.07	0.13	0.10 (0.02)	0.08	0.14	0.10 (0.01)	0.08	0.19	0.11 (0.03)	0.11	0.60	0.23 (0.12)
	1.5	0.11	0.20	0.15 (0.03)	0.12	0.20	0.15 (0.02)	0.12	0.28	0.17 (0.05)	0.17	0.90	0.35 (0.18)

Note. *Min* - minimum, *Max* - maximum, *M* - mean, *SD* - standard deviation of the mean.

reported from models using treatment coding.

To assess the significance of the fixed effects, we used a deviance table with Type III Wald Chi-square tests obtained with the *car* package (version 3.1–3; Fox & Weisberg, 2019) on the models fitted with sum coding to ensure the validity of main effects in the presence of interactions. All multiple comparisons for significant interactions were performed with Holm *p*-value correction with the *emmeans* package (version 1.10.6; Lenth, 2025). For all comparisons, the alpha level was set at 5%.

3.1. Appearance maps

To visually inspect individual responses to each target letter, eye, size and crowding condition, each observer's responses were plotted together (Figs. S3-S7 in the *Supplementary Materials)*. Further, to visually compare the average appearance of each target letter for each group, eye, crowding and size condition, all observers' response images were averaged pixel-by-pixel (Fig. 2). Specifically, the gray values of the corresponding pixels on each response image (from a particular group, target letter, eye, crowding and size condition) were summed with all other images and then divided by the total number of response images to yield an average pixel value.

3.2. Accuracy

Accuracy was calculated as the average of the proportion of hits (i.e., correctly selected squares on the grid) and the proportion of correct rejections (i.e., correctly not selected squares) yielding a proportion correct score between 0 and 1. Proportion correct scores were averaged across targets for each participant, eye, size and crowding condition. As the mean proportion-correct scores were bounded between 0 and 1 (accuracy range: 0.33–0.98), a beta distribution with a logit link function was chosen to model the accuracy data. To account for significant heterogeneity in the residual variance, the dispersion parameter (φ) was modelled as a function of the group and size (Group + Size), allowing the residual variance to differ across these conditions (Brooks et al.,

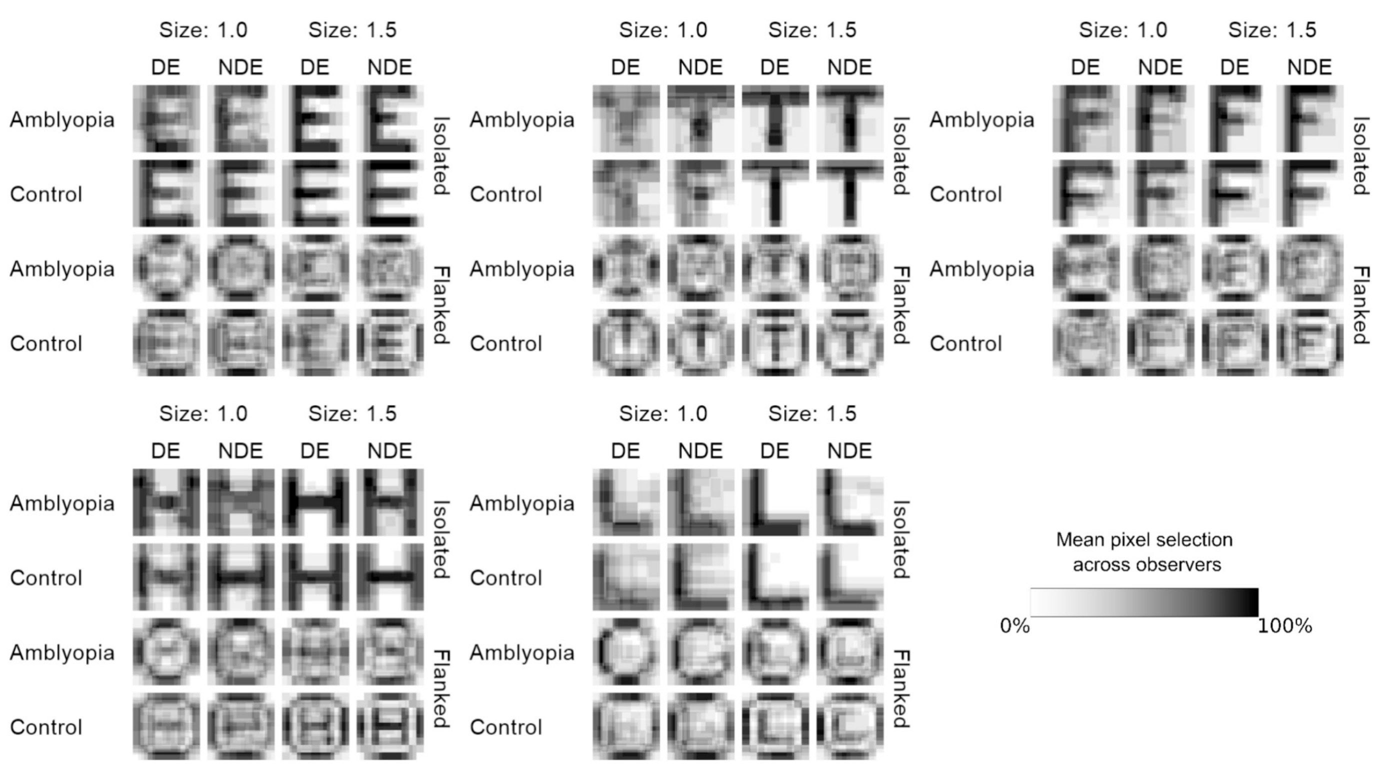

Fig. 2. Mean pixel selection in the responses of both groups (control, amblyopia) for each target letter in both sizes (1.0 and 1.5 x threshold), eye (DE/FE, NDE/AE) and crowding conditions (Isolated, Flanked). Please note that these appearance maps, which are averaged across observers, may obscure individual differences, especially amongst participants with amblyopia. Individual appearance maps are presented in the *Supplementary Materials* Figs. S3, S4, S5, S6, S7 for the E, T, F, H and L targets, respectively.

2017). The best supporting models found included a model with group, crowding condition and size as main effects (ProportionCorrect ~ Group + Crowding_condition + Size + (1|Participant_ID); AIC: −361.838) and a full model with all factors as main effects and an interaction between group and crowding condition (ProportionCorrect ~ Group * Crowding_condition + Size + Eye + EyeOrder + (1|Participant_ID); AIC: −361.84). There were no significant differences between the model with group, crowding condition and size as main effects and the one containing the interaction (all *ps* > 0.05). As the full model with the interaction had similar AIC and fulfilled model assumptions best it was chosen for inference.

For the chosen model, the overall distribution of the standardized residuals was checked (see *3. Analysis* section above) and the QQ plot of the expected vs observed plot as well as the Kolmogorov-Smirnov test ($p = .05$) indicated that the beta distribution was appropriate. No clear patterns were observed in the fitted vs residuals plot and the quantile test was not significant (location of quantiles test: $p = .17$) indicating that the linearity assumption was upheld. No significant dispersion was found (dispersion = 0.95, $p = .63$). Residuals were also plotted across the levels of each categorical predictor as mentioned above (section *3. Analysis*) and all Kolmogorov-Smirnov tests (to check for deviation from uniformity) and Levene tests (to check for homogeneity of variance) were not significant (all *ps* > 0.05), except for the group factor, where deviations from uniformity in amblyopia were detected (KS: $p = .03$) and the Levene test was significant ($p = .03$). A significant Levene test for the group factor indicated different variances between the control and amblyopia group, which can be attributable to the inherent differences already in both populations. As all other tests and the quantiles showed no significant deviations, the model was still used for inference. One outlier was found, but not found to be significant in the outlier test ($p = 1$). No collinearity of the factors was present. Random effects were normally distributed (Shapiro-Wilk test: $p = .17$).

3.3. Response-target structural similarity (SSIM)

To assess the similarity between the response shapes and the presented stimulus, each response on the grid was saved as an image and compared to the target with the structural similarity index (SSIM; Wang et al., 2004). The SSIM compares corresponding spatial patches/sliding windows from two images in regard to structure, luminance, and contrast, resulting in a single mean structural similarity value for each image pair. The structure comparison is based on the inner product of vectors of the images' normalized signal (Wang et al., 2004). The output of the SSIM comparison is a similarity map, i.e., an array containing the local similarity values, as well as a numeric scalar SSIM score. The numeric scalar SSIM score was used here for analysis. The SSIM score ranges from −1 (anti-correlation) to 1 (highest correlation; images are identical), with 0 representing no correlation between the images. To ensure that we were able to directly compare each response's shape to the corresponding stimulus shape, the response images (740 responses, approximately 64%) were scaled and centered prior to the SSIM comparison. Specifically, the response images were scaled up or down until either the width and/or the height of each response shape was the same size as an expected correct response on the grid. Scaling was performed with the nearest-neighbor interpolation of the *imresize* function (MATLAB R2022a, 2022; The MathWorks Inc., 2022). The structural similarity index is typically used with complex images to assess image quality based on the degradation of structural information, such as due to added blur or image compression (Wang et al., 2004). To evaluate to what extent accuracy and the structural similarity scores were related, a two-sided Spearman's rank correlation between the two measures was performed (see Fig. S1 in the *Supplementary Materials*). As the structural similarity scores are bounded (between −1 and 1), we transformed the SSIM scores into a 0–1 range by adding 1 and dividing by 2 (i.e., SSIM' = (SSIM +1) / 2)). This allowed us to model the SSIM scores with other families than a Gaussian within the exponential dispersion model (EDM) family. The transformed structural similarity (SSIM) scores were averaged across targets for each participant, eye, size and crowding condition.

A beta distribution with a logit link function was found to best model the data. To account for observed heteroscedasticity between groups, dispersion was modelled as a function of fixed effects (Brooks et al., 2017) (~ Group). Two best-supported models were found: a *reduced model* including only the group, crowding condition and size as fixed effects (SSIM' ~ Group + Crowding_condition + Size + (1|Participant_ID), AIC: −706.01) and a *full model* with all independent variables and an added interaction between group and crowding condition (SSIM' ~ Group * Crowding_condition + Size + Eye + EyeOrder + (1|Participant_ID); AIC: −705.08). The models did not differ significantly ($X^2(3) = 5.07, p = .17$). Further, as the AIC values did not differ by more than 1 unit, the *full model* was chosen for inference as it contained the interaction of primary interest in our study (Group * Crowding condition). Model adequacy was confirmed by visual inspection of the expected and observed residual distribution and a Kolmogorov-Smirnov test ($p = .41$ for the model chosen for inference), indicating no significant deviation from a beta distribution. Linearity was checked by plotting the fitted values against the residuals and no apparent patterns were observed. The quantile test was also not significant (location of quantiles test: $p = .11$), further indicating that the linearity assumption was upheld. Residuals were also plotted across the levels of each categorical predictor as mentioned above (section *3. Analysis*) and all tests (to check for deviation from uniformity and homogeneity of variance) were not significant (all *ps* > 0.05). No influential outliers were found ($p = .71$). The variance inflation factor (to check for collinearity of the factors) was always lower than 5. Dispersion was not significant (dispersion = 0.94, $p = .66$). The random effects were normally distributed (Shapiro-Wilk test: $p = .67$).

3.4. Perceived similarity (PSS)

SSIMs do not always capture perceived similarity between shapes. To obtain perceived similarity scores (PSS), human observers rated the similarity between targets and responses. First, we obtained all unique responses by having two raters (one of them the first author) separately compare each response to all other responses and sort them into two groups: same vs. not same (Fig. S2A of *Supplementary Materials*). For images categorized as identical, comparisons with the other images were only made once. Agreement between the two raters was calculated based on all possible comparisons for all targets (general Cohen's kappa: 0.98; general percent agreement: 99.99%), as well as separately for isolated targets (Cohen's kappa for the E: 0.93, T: 1.00, F: 0.99, H: 0.99, L: 0.94; percent agreement for the E: 99.97%, T: 100.00%, F: 100.00%, H: 99.99%, L: 99.97%), and for flanked targets (Cohen's kappa was 1.00 for each individual target separately; percent agreement was 100% for each target separately**)**. There were 965 images that were unique responses (365 for isolated targets and 600 for flanked targets) based on the sorting from the first author. Four of the 965 response images were identical to the corresponding targets and one was an empty response so they were not included in the similarity rating task, yielding 960 unique responses that were used in the following similarity rating task to obtain perceived similarity scores.

In the similarity rating task, five naive observers (who did not perform the main experiment) were presented with image pairs consisting of one of the presented targets in the main experiment and a unique response to that target. Six pairs were presented simultaneously in two columns on the left side of the screen (Fig. S2B of *Supplementary Materials*; as the total number of comparisons was not divisible by six, some trials contained fewer than six comparisons). Each image subtended 2.55° by 2.55° of visual angle. The target images were presented in the leftmost column and the unique responses to the right of the targets. To the right of each image pair, a rating scale from 0 to 100 was presented. The scale showed tick marks every 10 units with the halfway mark labelled '50'. The labels on the extremes of the rating scale were counterbalanced across observers: half of observers rated from 0 (Very different) to 100 (Very similar); the other half rated from 0 (Very similar) to 100 (Very different). Raters were instructed to rate the similarity of each image pair based on the shape, including proportions (e.g., a T response with a wider stem than the target T was rated as less similar than the same T response with the same width of the stem as the target T). The rating task was blocked by target x crowding condition: 10 blocks (5 target letters (E, T, F, H, L) x 2 crowding conditions (isolated, crowded)). Block order and target-response comparisons included in each trial were randomized. The similarity results from the rating task were averaged across raters for each response-target pair. These response-target similarity scores were averaged across targets for each participant, eye, size and crowding condition to yield 'perceived similarity' scores (PSS). An initial inspection of PSS revealed a bimodal distribution characterized by distinct peaks corresponding to the isolated and flanked experimental conditions. Further, as the PSS were bounded between 0 and 100 (range 12.07–89.45), we transformed the PSS into a 0–1 range by dividing all scores by 100 which allowed us to use other distributions than a gaussian within the exponential dispersion model (EDM) family. A beta distribution with a logit link function was found to best model the data. As in the previous analyses, to account for significant heterogeneity in the residual variance, the dispersion parameter (φ) was best modelled as a function of the group and crowding condition (Group * Crowding_condition) (Brooks et al., 2017).

Three best-supporting models were compared: a *reduced model* including only the group, crowding condition, size and eye as a fixed-effects factor (PSS ~ Group + Crowding_condition + Size + Eye + (1|Participant_ID); AIC: −306.37), a model including also eye order (PSS ~ Group + Crowding_condition + Size + Eye + EyeOrder + (1|Participant_ID); AIC: −304.43), and the *full model* with an added interaction between group and crowding condition (PSS ~ Group * Crowding_condition + Size + Eye + EyeOrder + (1|Participant_ID); AIC: −303.28). Although adding eye order, and the group x crowding interaction did not significantly improve the model (all $ps > 0.05$) (compared to a model with just group, crowding condition, size and eye as main effects), we retained them to report the significance of these effects. Further, although the model with the group x crowding condition interaction did not have the lowest AIC, it was chosen for inference because this interaction was the main focus of the study. For the chosen model, the overall distribution of the standardized residuals was uniform as checked by plotting the expected distribution of the residuals against the observed distribution from the model (Kolmogorov-Smirnov test; $p =$ .38.) Linearity was checked by plotting the fitted values against the residuals. Dispersion was also checked (dispersion = 1.06, $p = .58$). The residuals were also plotted across the levels of each categorical predictor (all $ps > 0.05$ for Kolmogorov-Smirnov tests and Levene tests, except for the crowding condition, i.e., flanked $p = .003$ and $p = .01$, respectively). Residual deviations in the crowding conditions are attributable to the inherent difference in response variance between isolated and flanked stimuli, which persists even after explicitly modelling variance as a function of group and crowding condition. Visual inspection of score densities and observed-vs-predicted plots confirmed significant response compression in the flanked condition, particularly at the 0.25 quantile of the response distribution. 25% of the perceived similarity scores fall below a PSS of 0.40 in the isolated condition, and below of a PSS of 0.21 for the flanked condition. As all other quantiles showed no significant deviation in the residuals vs. predicted plot and the variance between crowding conditions is theoretically expected, the model was still used for inference. The outlier test was not significant ($p = .12$). The variance inflation factor (to check for collinearity of the factors) was always lower than 5. The random effects were normally distributed (Shapiro-Wilk test: $p = .47$).

4. Results

We asked observers to report the appearance of isolated and flanked letters presented foveally to the dominant/fellow and non-dominant/amblyopic eye using a novel method to capture appearance. Fig. 2 shows the average appearance maps of both groups. Our main finding was that accuracy, structural and perceived similarity were lower for observers with amblyopia compared to controls, for small compared to large stimuli, and for flanked compared to isolated targets. Importantly, for accuracy and SSIM (though not for perceived similarity, PSS), there was an interaction between Group and Crowding, with stronger crowding for amblyopes compared to controls.

4.1. Appearance maps

Individual appearance maps, i.e., each observer's response to all target letters at each eye, size and crowding condition are shown in Figs. S3-S7 in the *Supplementary Materials.* Individual maps revealed distortions in target appearance, i.e., observers perceived (and depicted) target elements differently than presented. Similar 'error categories' that were used to quantify variability in target appearance in previous papers on isolated and flanked targets (Sayim & Wagemans, 2017) can be observed in the individual appearance maps of our study (Fig. S3-S7 in the *Supplementary Materials*). For example, we found truncations of elements (i.e., depicting elements shorter than they were in the target; e.g., AMBFOV_3's responses to the E target), and extension of elements (e.g., for target E, AMBFOV_16, 22 and 28 depicted the inner horizontal element of the E longer than it was presented). Often, flankers were also depicted as connected to rather than separate from each other and the target (e.g., AMBFOV_40 responses to target E, Fig. S1 in the *Supplementary Materials*). Shape distortions with straight lines being depicted as curved (as far as curvature can be depicted with limited numbers of squares) were also observed (e.g., responses to the E and F targets made by AMBFOV_41). These distortions (curved lines) are also visible in the averaged data as can be seen in Fig. 2 for the crowded conditions (especially for targets T and L).

Importantly, such distortions in target appearance impact accuracy, structural (SSIM) and perceptual similarity (PSS) scores, however, we did not quantify which specific distortion yielded what reduction of accuracy, structural similarity and perceived similarity. Distortions in target appearance can be observed both in the individual (Figs. S3-S7 in the *Supplementary Materials*) and averaged appearance responses (Fig. 2).

4.2. Accuracy

Using the best model found to analyze accuracy (Table S2 in the *Supplementary Materials*; ProportionCorrect ~ Group * Crowding_condition + Size + Eye + EyeOrder + (1 | Participant_ID); AIC = −361.84), accuracy (proportion correct, Fig. 3) was higher for observers in the control group ($M = 0.66$, $SEM = 0.01$) compared to the amblyopia group ($M = 0.59$, $SEM = 0.01$) ($X^2(1) = 12.95$, $p < .001$), at 1.5 x threshold ($M = 0.66$, $SEM = 0.01$) compared to 1.0 x threshold size ($M = 0.59$, $SEM = 0.01$) ($X^2(1) = 20.64$, $p < .001$), and for isolated ($M = 0.67$, $SEM = 0.01$) compared to flanked targets ($M = 0.58$, $SEM = 0.01$) ($X^2(1) = 50.52$, $p < .001$) (see Fig. 4 for the crowding strength for each group, eye and size combination). No other main effects were found.

An interaction between group and crowding condition was significant ($X^2(1) = 4.76$, $p = .03$), indicating that the effect of crowding on accuracy differed between controls and observers with amblyopia. No other interaction was significant. Post-hoc analysis corrected for multiple comparisons revealed that accuracy was higher for isolated compared to flanked targets for controls (isolated: $M = 0.69$, $SEM = 0.02$; flanked: $M = 0.63$, $SEM = 0.02$; z-ratio = 2.97, $p = .009$), and for amblyopes (isolated: $M = 0.65$, SEM = 0.01; flanked: $M = 0.52$, $SEM = 0.01$; z-ratio = 8.29, $p < .001$), revealing a crowding effect for both

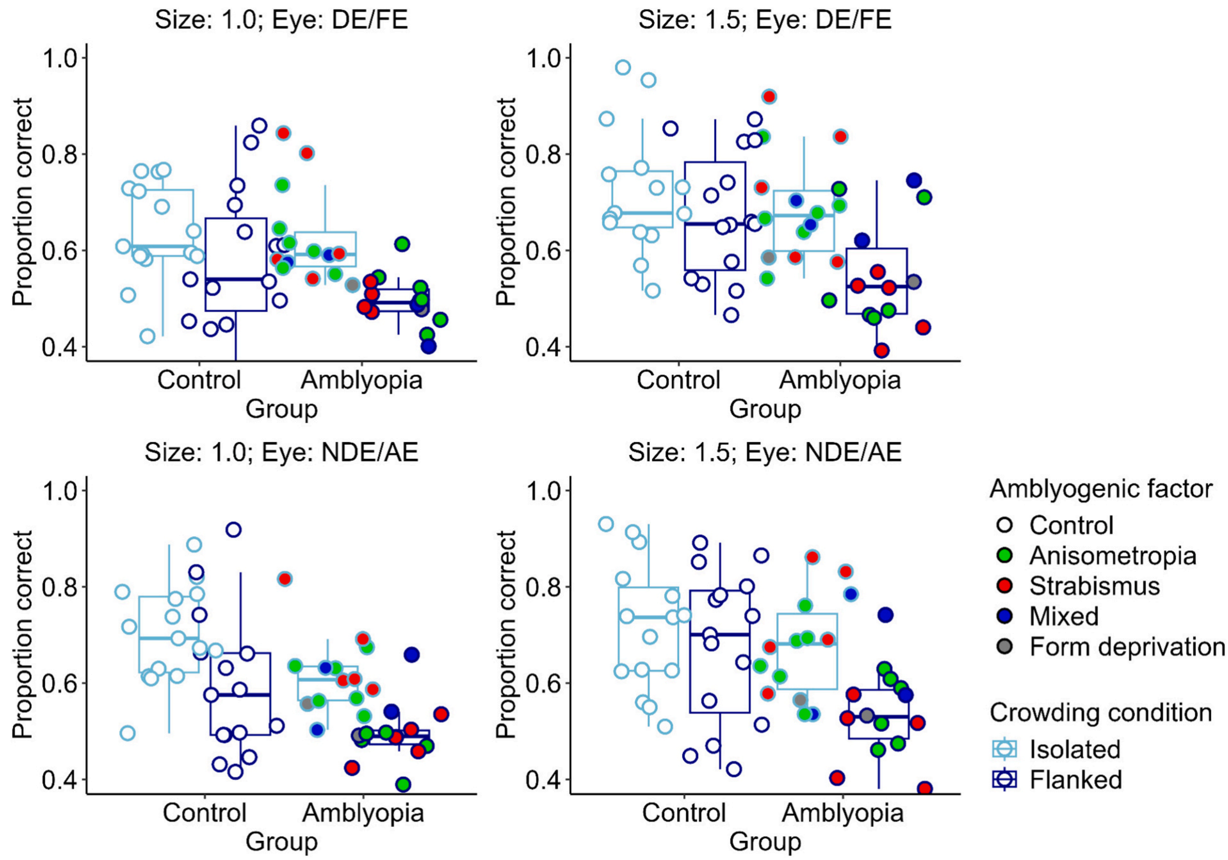

Fig. 3. Proportion correct for both groups and crowding conditions, at small (1.0; left) and large (1.5 x threshold; right) size for each eye (top: dominant/fellow eye, bottom: non-dominant/amblyopic eye). Each datapoint represents a single observer. Data point color represents the underlying amblyopic factor (green: anisometropia; red: strabismus; blue: both anisometropia and strabismus; black: form deprivation). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

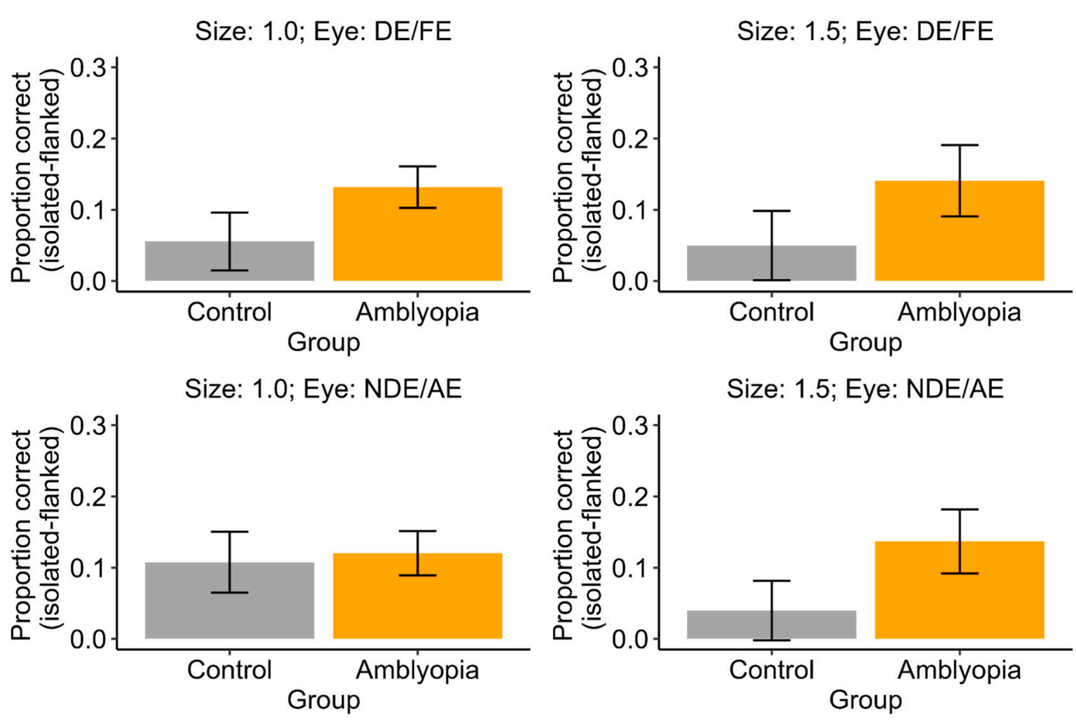

Fig. 4. Crowding strength for accuracy of each group at 1.0 (left) and 1.5 times threshold (right) for both the dominant/fellow eye (top) and non-dominant/amblyopic eye (bottom). Error bars represent the standard error of the mean.

groups. Importantly, the effect of flankers was larger in observers with amblyopia compared to controls (z-ratio = 4.26, $p < .001$) showing stronger crowding in amblyopia. (Reported for completeness: Accuracy was also higher for controls with isolated targets compared to observers with amblyopia with flanked targets (z-ratio: 6.79, $p < .001$)). No other comparison was significant.

4.3. Response-target structural similarity (SSIM)

Using the full SSIM model (Table S3 in the *Supplementary Materials*), response-target structural similarity (Figs. 5 and 6) was highest for controls ($M = 0.79$, $SEM = 0.01$) compared to amblyopes ($M = 0.76$, $SEM = 0.005$; $X^2(1) = 6.16$, $p = .01$), when targets were presented at 1.5 x threshold ($M = 0.79$, $SEM = 0.006$) compared to 1.0 x threshold size ($M = 0.76$, $SEM = 0.005$) ($X^2(1) = 33.44$, $p < .001$), and when targets were isolated ($M = 0.78$, $SEM = 0.006$) compared to when flanked ($M = 0.76$, $SEM = 0.006$, $X^2(1) = 7.52$, $p = .01$) (Figs. 5 and 6). No other main effects were found.

An interaction between group and crowding condition was significant ($X^2(1) = 4.31$, $p = .04$), indicating that the effect of crowding on SSIM differed between groups. No other interactions were found. Post-hoc analysis corrected for multiple comparisons revealed significantly higher structural similarity in observers with amblyopia for the isolated ($M = 0.77$, $SEM = 0.008$) compared to the flanked condition ($M = 0.74$, $SEM = 0.005$; *z-ratio* = 4.69, $p < .001$), but not for controls (isolated: $M = 0.79$, $SEM = 0.01$, flanked: $M = 0.79$, $SEM = 0.01$, *z-ratio* = 0.39, $p = .70$), revealing a crowding effect only in amblyopia. In the flanked condition, similarity was higher for controls ($M = 0.79$, $SEM = 0.01$) compared to amblyopes (*z-ratio* = 3.12, $p = .007$), as well as for controls in the isolated condition ($M = 0.79$, $SEM = 0.008$) compared to amblyopes in the flanked condition (*z-ratio* = 3.38, $p = .004$). No other comparison was significant.

4.4. Perceived similarity (PSS)

Using the full model with the group x crowding condition interaction (Table S4 in the *Supplementary Materials*; PSS ~ Group * Crowding_condition + Size + Eye + EyeOrder + (1|Participant_ID); AIC: −303.28), perceived response-target similarity (Figs. 7 and 8) was higher for controls ($M = 0.45$, $SEM = 0.02$) compared to amblyopes ($M = 0.38$, $SEM = 0.02$, $X^2(1) = 4.47$, $p = .03$), and for isolated ($M = 0.52$, $SEM = 0.02$) compared to flanked ($M = 0.31$, $SEM = 0.01$) targets ($X^2(1) = 176.98$, $p < .001$). In addition, similarity was higher for 1.5 x threshold size ($M = 0.48$, $SEM = 0.02$) compared to 1.0 x threshold size ($M = 0.35$, $SEM = 0.02$) ($X^2(1) = 64.03$, $p < .001$). No other main effects or interactions were found.

5. Discussion

Observers with amblyopia have been reported to often perceive stimuli as distorted, both when the stimuli are presented in isolation (Barrett et al., 2003; Bradley & Freeman, 1985; Hess et al., 1978; Sireteanu et al., 1993) and with flankers (i.e., crowded; Pugh, 1958). Foveal crowding in amblyopia has been shown to be stronger (e.g., Hess & Jacobs, 1979) and to extend over larger distances than in controls (e.g., Hariharan et al., 2005; Levi, Hariharan, et al., 2002). Most previous studies on appearance in amblyopia (e.g., Barrett et al., 2003; Bradley & Freeman, 1985; Hess et al., 1978; Pugh, 1958) did not include a control

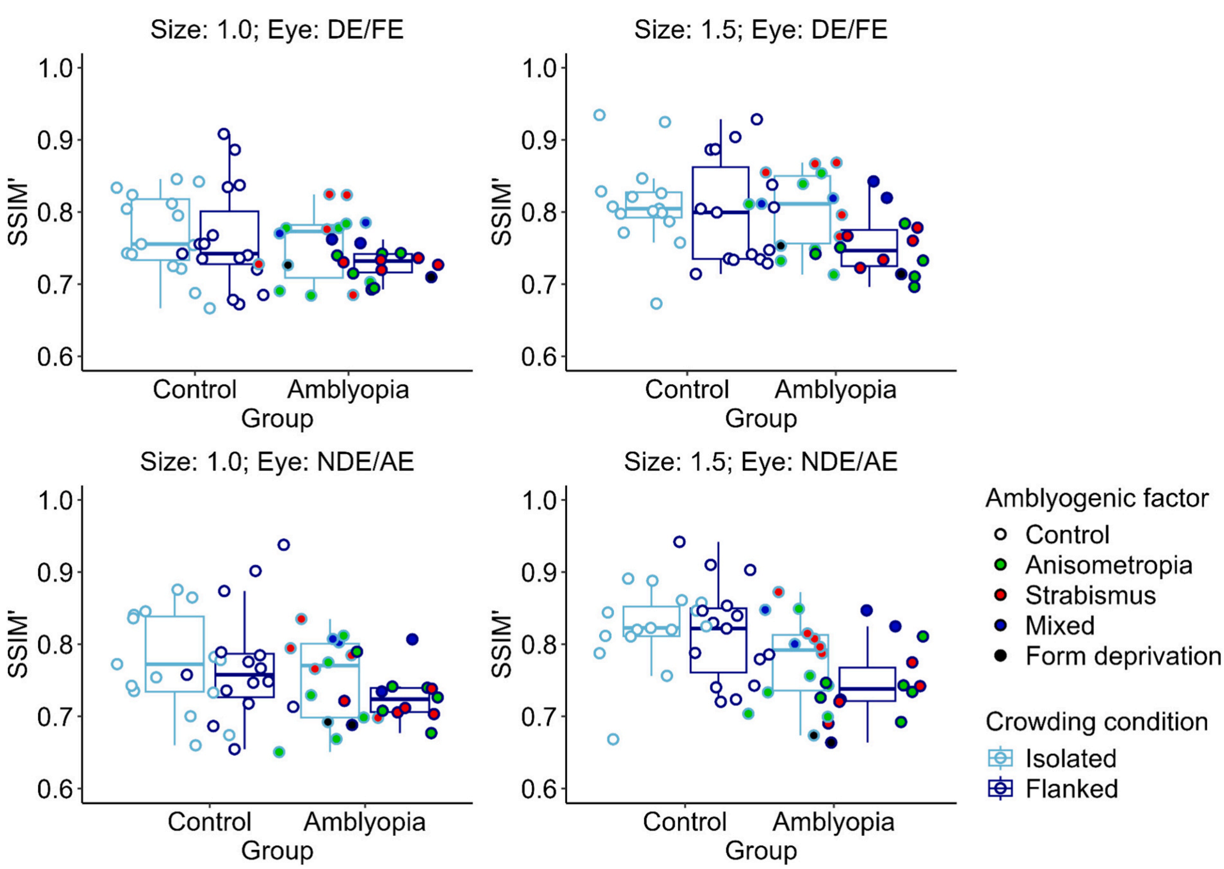

Fig. 5. Response-target structural similarity (SSIM') for both groups and crowding conditions, at 1.0 (left) and 1.5 (right) x threshold for each eye (top: dominant/fellow eye, bottom: non-dominant/amblyopic eye). Each data point represents a single observer. Data point color represents the underlying amblyopic factor (green: anisometropia; red: strabismus; blue: both anisometropia and strabismus; black: form deprivation). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

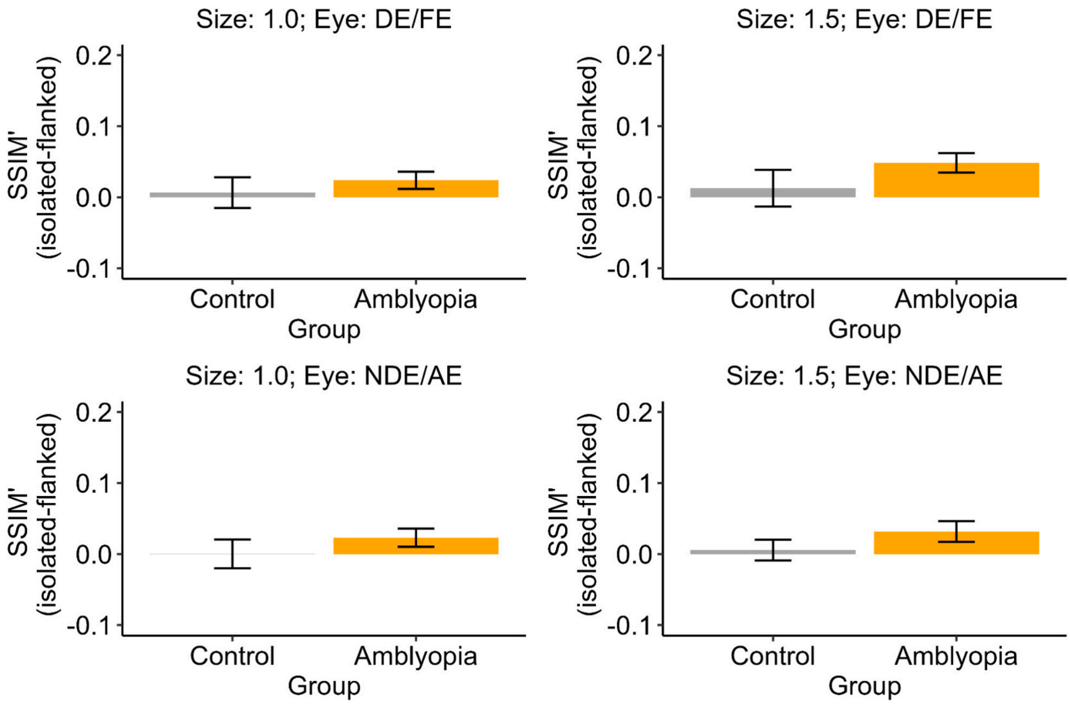

Fig. 6. Crowding strength for response-target structural similarity (SSIM') of each group at 1.0 (left) and 1.5 times threshold (right) for both the dominant/fellow eye (top) and non-dominant/amblyopic eye (bottom). Error bars represent the standard error of the mean.

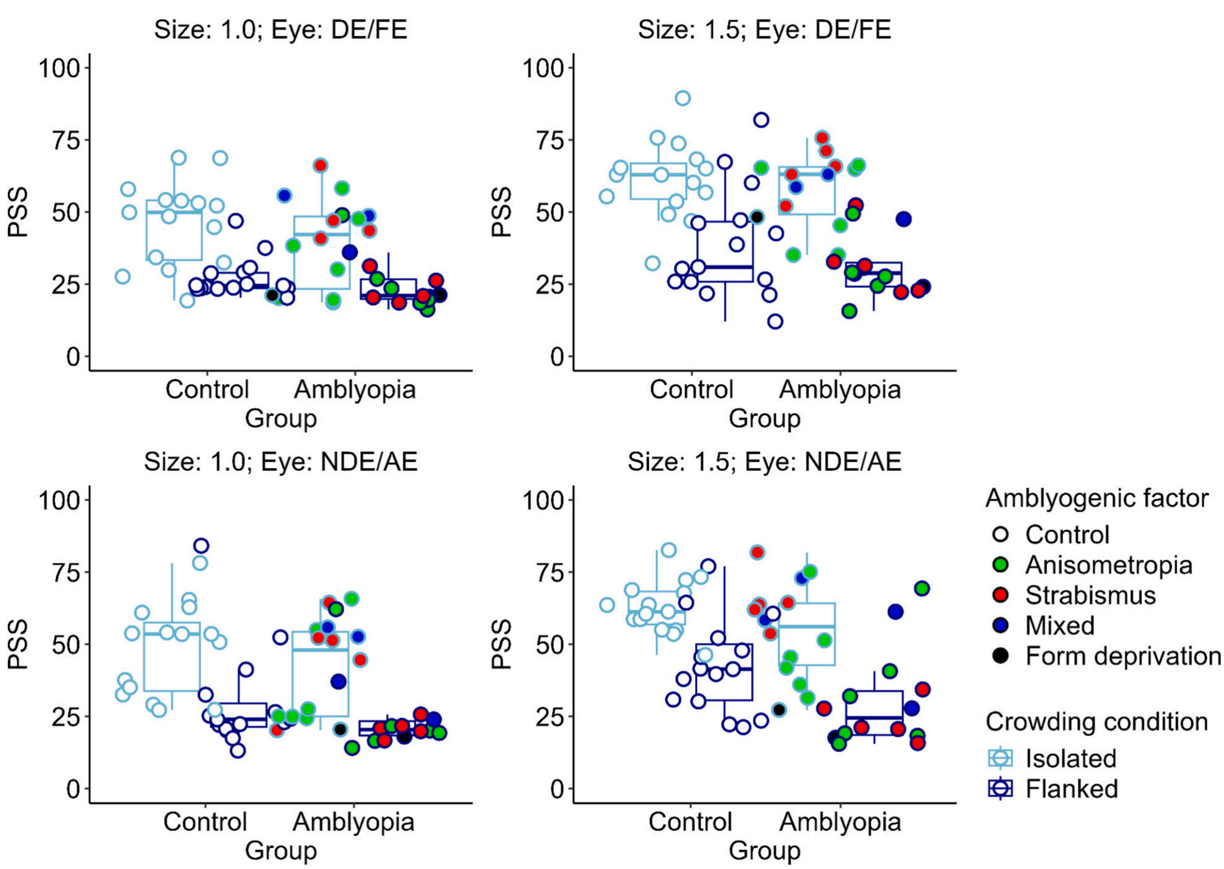

Fig. 7. Perceived target-response similarity (PSS) for both groups and crowding conditions, at 1.0 (left) and 1.5 (right) x threshold for each eye (top: dominant/fellow eye, bottom: non-dominant/amblyopic eye). Each data point represents a single observer. Data point color represents the underlying amblyopic factor (green: anisometropia; red: strabismus; blue: both anisometropia and strabismus; black: form deprivation). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

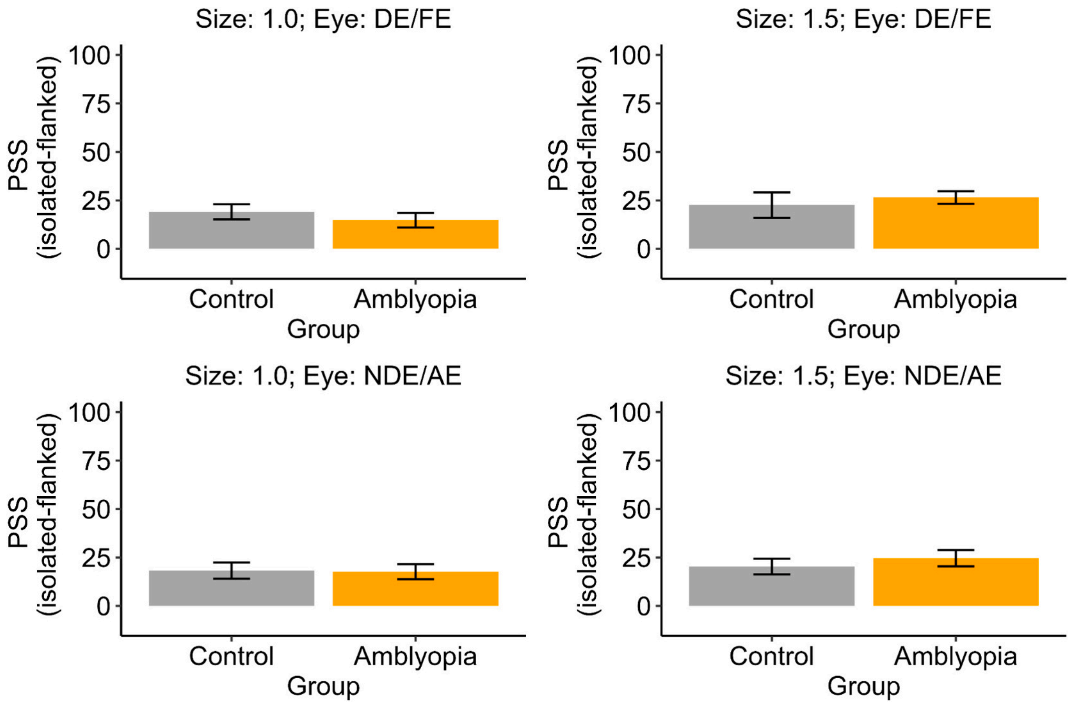

Fig. 8. Crowding strength for perceived target-response similarity (PSS) for each group at 1.0 (left) and 1.5 times threshold (right) for both the dominant/fellow eye (top) and non-dominant/amblyopic eye (bottom). Error bars represent the standard error of the mean.

group, and did not adjust stimulus size to compensate for acuity differences. In the current study, we investigated stimulus appearance of isolated and crowded targets in amblyopia and controls while individually scaling stimulus sizes to each observer's resolution threshold to control for acuity differences. Observers were presented with small and large target letters (1.0× and 1.5× their resolution threshold, respectively) to the amblyopic/non-dominant eye or fellow/dominant eye. Target letters were either presented in isolation or flanked by four bars. Observers' task was to report target appearance using a 9 × 9 square grid. We calculated accuracy, response-target similarity (SSIM; Wang et al., 2004), and perceived response-target similarity (PSS).

We found that overall accuracy, structural similarity of responses, and perceived response-target similarity of observers with amblyopia were lower than of controls, despite the individually adjusted stimulus sizes. As expected, all three measures were lower in the flanked compared to the isolated condition. In the crowding condition, accuracy and structural similarity were higher for controls than for observers with amblyopia, whereas perceived response-target similarity did not differ between groups. Comparing the two stimulus sizes revealed that accuracy was higher with the large (1.5 x resolution threshold) compared to the small stimuli (1.0 x resolution threshold). Large stimuli also yielded responses that were more similar to the target compared to small stimuli (as shown by the SSIM and PSS metrics).

5.1. Observers with amblyopia and controls

Overall, accuracy, response-target structural similarity (SSIM) and perceived similarity (PSS) scores were lower for observers with amblyopia compared to controls, across all eye, crowding, and size conditions, despite the individually adjusted stimulus sizes. These results suggest that visual deficits in amblyopia cannot be accounted for by reduced acuity alone, and may reflect a combination of different deficits, possibly including decreased contrast sensitivity (Levi & Harwerth, 1980; Zele et al., 2007), increased internal noise (Acar et al., 2019; Levi & Klein, 2003; Pelli et al., 2004; Wang et al., 2017), and decreased position acuity (Levi, 2021). For example, Levi and Klein (1983) asked observers with amblyopia to perform a vernier alignment and a bisection task. They found impaired position acuity in observers with amblyopia, especially strabismic and mixed amblyopia, and showed that these spatial localization deficits were not accounted for by resolution deficits, unlike in anisometropic observers. In our sample, most of the observers with amblyopia had either mixed or strabismic amblyopia. Hence, one explanation for the relatively low accuracy found for observers with amblyopia may be their decreased position acuity impacting the perceived relative position of the target features. Cortical differences may also explain the differences found here. Previous research has found increased cortical noise (i.e., neural response variability) in V1 (Acar et al., 2019) and V2 (Wang et al., 2017), as well as increased multiplicative internal noise in amblyopia (Levi & Klein, 2003; Pelli et al., 2004).

Closely examining individual observers' responses revealed significant inter-individual differences in both groups. Some of the responses of observers with amblyopia and controls showed few or no distortions in target appearance while others revealed strong distortions (Figs. S3-S7 in the *Supplementary Materials*). These results are similar to the findings by Barrett et al. (2003). In their study, observers were presented with gratings varying in spatial frequency, orientation and contrast, and asked to indicate if there were perceptual differences when seeing the stimulus monocularly with their fellow vs. amblyopic eye. When perceptual differences were reported, observers sketched the appearance of the grating stimuli as seen by the amblyopic eye. They found two different observer groups: one who reported distortions in presented (undistorted) gratings (e.g., grating lines were perceived as wavy, and/or with abrupt offsets; 'non-veridical' group), and another group who did not report any distortions ('veridical' group). In both groups there were observers with good (<1 logMAR) and poor acuity in the amblyopic eye (>1 logMAR), suggesting that the distortions did not depend on the depth of amblyopia. Hence, our results and those of Barrett et al. (2003) suggest that perceived target distortions are foremost a by-product of cortical processing (both for controls and observers with amblyopia). Other factors such as high order optical aberrations may also play a role in target distortions; however, prior research that bypassed the optics of the eye with laser interferometry (e.g., Sharma et al., 1999 in amblyopia; Williams, 1985), also found distortions in target appearance. Taken together, these results are consistent with the view that cortical processing underlies the distortions found in target appearance in both observers with and without amblyopia.

A potential factor contributing to the group differences in our study is age. Although we attempted to age-match both groups, the amblyopia group included three older adults over 40 years of age. Previous studies have shown aging effects in crowding, in particular at more eccentric locations, where older adults required larger stimuli to reach a resolution threshold under crowded conditions in the parafovea and in the periphery (Scialfa et al., 2013). However, when taking the ratio of crowded to uncrowded acuity, age differences did not persist (Scialfa et al., 2013). In the fovea, no differences were found between younger and older adults (mean age of 60 years old; Shamsi et al., 2022). More recently, Rappeport et al. (2025) investigated foveal crowding in both young and older adults (20–89 years old), and found significant differences between younger (<50 years old) and older adults (>50 years old), but only for brief presentation durations (120 ms). Since our study used longer presentation times (500 ms), it remains unclear whether age would have an effect and how much it might affect our findings. Nevertheless, we acknowledge that age could play a minor role in the observed group differences. Thus, to test if age was impacting our findings, we included age as a factor in all analyses (see *Supplementary Materials*), however age was not a significant predictor in any of the analyses.

The method used in the current study captures target appearance with greater detail compared to other tasks (e.g., standard alternative-forced-choice), which typically focus on one or two target dimensions/features (e.g., orientation). In contrast to letter identification tasks, each presented square of the target and the background could be reported correctly or incorrectly. Hence, accuracy measures are more fine-grained than in typical identification tasks. Consequently, a low accuracy score does not necessarily imply errors in a standard letter identification task: shifts, truncations, or omissions of elements –each composed of squares– may still allow correct letter identification, while lowering accuracy scores as the responses still deviate from the presented target in our study. Moreover, additional metrics can be simultaneously employed to investigate target appearance (e.g., similarity metrics such as the SSIM and (visual inspection of) appearance maps, *Supplementary Materials* Figs. S3–7). Beyond individual variability, the averaged appearance maps (Fig. 2) provide a visual illustration of the variability of responses and thereby the observed effects. For isolated targets, especially at 1.5 x threshold size, controls showed highly consistent responses closely matching the presented letters. However, because accuracy was calculated at the square ("pixel") level, even small deviations reduced accuracy values despite the identifiable overall shape. In contrast, crowding strongly increased the variability of responses, producing maps in which the target letter is sometimes hardly –or not at all- identifiable. Comparisons across groups are also reflected in the differences between appearance maps, with more variable responses in conditions where accuracy was low and the other appearance measures yielded low levels of response-target similarity. Similarly, the effect of target size is visually represented in the appearance maps by showing the increase of target visibility with larger sizes. Importantly, appearance maps, both for individuals and averaged, provide a representation to compare responses in regard to qualitative differences and derive hypotheses about the possibly systematic deviations and distortions of target appearance and their underlying mechanisms in individuals and observer groups. Beyond providing a more detailed account of the observer's visual experience, the current method is faster and more precise than, for example, drawings. This efficiency is particularly important when presenting stimuli in difficult viewing conditions (e.g., near the resolution limit, under crowding) or with populations that are visually impaired, such as amblyopes, in order to minimize the experimental time, fatigue, and frustration.

5.2. Crowding

Crowding effects were observed in both groups: accuracy, SSIM and PSS were lower for flanked compared to isolated letters (Figs. 4, 6 and 8). Foveal crowding has been shown previously in observers with amblyopia (e.g., Bonneh et al., 2004, 2007; Polat et al., 2005; Pugh, 1958; Song et al., 2014) and controls (e.g., Danilova & Bondarko, 2007; Sayim et al., 2008, 2010; Malania et al., 2007; Coates et al., 2018). Importantly, there was a significant interaction between group and crowding for accuracy and SSIM: crowding was stronger in amblyopes compared to controls. These results are in line with previous findings showing stronger crowding in observers with amblyopia compared to controls across a range of stimuli (e.g., Hess & Jacobs, 1979; Song et al., 2014). By contrast, no significant group x crowding interaction was found with the perceived similarity scores (PSS). Differences between accuracy and SSIM on the one hand and PSS on the other likely reflect the levels at which these measures operate. Accuracy operates on the pixel-level, quantifying correct and incorrect pixels/grid squares for each response. SSIM is an image-based metric that evaluates local structural shape similarity and –unlike pixel-based accuracy– captures fine-grained, deviations, attempting to mimic the functionality of the visual system. Both measures are sensitive to small distortions (or deviations from similarity), and thereby relatively well suited to capture group differences. By contrast, PSS is likely less sensitive to smaller pixel-level deviations. Instead, PSS appears to be strongly influenced by the recognizability or categorization of shapes: for example, two responses that are both categorized as the same letter may be rated as more similar than two responses that differ to the same extent at the pixel level, but are not categorized as the same letter. Hence, the absence of a group x crowding interaction in PSS may indicate that, while amblyopes' responses showed reduced accuracy (i.e., a lower proportion of correctly selected or correctly rejected squares on the grid) and structural deviations under crowding (as measured by SSIM), these deviations did not necessarily reflect perceptually salient differences for human observers rating the responses in regards to response-target similarity.

Importantly, the Group x Crowding interaction found with the SSIM seems to be mostly driven by significant crowding in amblyopia. Controls showed no difference in SSIM between isolated and flanked conditions, indicating no crowding effect in controls (Fig. 6) despite presenting stimuli at sizes near the resolution limit to ensure crowding could be measured in both groups (Danilova & Bondarko, 2007). Moreover, the absence of significant differences in structural similarity (SSIM) between controls and amblyopes in the isolated condition may suggest that observers with amblyopia did not perceive stimuli as more distorted than controls in the absence of crowding. Taken together with the PSS analysis results, which revealed a main effect of Group but no Group x Crowding interaction, these findings suggest that observers with amblyopia may perceive stimuli more distorted compared to controls even when the visual acuity deficit is taken into account. Under crowding, perceived similarity may have been less affected, likely because many targets could still be identified by the raters.

5.3. Eye and eye order

No significant differences in accuracy, SSIM or PSS were found between the amblyopic/non-dominant and the fellow/dominant eye, or as a function of the eye order, in observers with amblyopia and controls. For isolated targets the absence of an eye effect was expected because the target size was scaled to each individual eyes' resolution threshold, equating target visibility for both eyes. Acuity and sensitivity can be worse in both eyes of observers with amblyopia, i.e., also in the fellow eye, (Kandel et al., 1980); this scaling therefore ensured that target resolution was approximately equal in both eyes. Size scaling does not rule out that crowding still differs between the two eyes. Eye differences could have been found both in observers with amblyopia and controls because crowding does not scale linearly with acuity. Moreover, some amblyopic deficits, such as positional uncertainty, may not necessarily scale with stimulus size. Nevertheless, no differences between the amblyopic and the fellow eye or between the non-dominant and dominant eye in controls, were observed, indicating that crowding strength did not differ between the eyes. These results are consistent with Norgett and Siderov (2017) in which no stronger crowding for the amblyopic compared to the fellow eye has been found.

We also examined whether the order in which the eyes were tested influenced any of our measures (accuracy, SSIM, PSS). Eye order could have influenced responses as viewing stimuli with one eye first may provide prior knowledge about a target that could influence responses when the same target is subsequently viewed with the other eye. However, no effects of eye order were observed for accuracy, SSIM, or PSS indicating that any potential influence of target knowledge did not influence performance or appearance.

5.4. Target size

As expected, accuracy, and both response-target similarity measures (SSIM and PSS) were significantly lower for small compared to large stimuli. Testing different sizes enabled us to both investigate how performance and appearance depend on retinal size, and served as a baseline to check if observers could use the grid successfully.

The effect of size is in line with a previous study in which we presented symbols at 0.5, 0.67, 1.0 and 1.5 times the observers' resolution threshold, using the same paradigm to capture appearance (Gomes Tomaz et al., 2022). In that study, observers with normal vision reported target appearance on a 5 × 5 square grid. The results showed clear differences in reported appearance between the smallest sizes and sizes at threshold, as well as between threshold sizes and sizes at 1.5 times the threshold. The effect of size observed in the present study was therefore expected, and it shows that the method to capture appearance used here is sensitive enough to capture both accuracy and appearance differences. Importantly, the advantage in the large size condition confirms that potential confounding factors, such as motor skills and ability to use the grid interface, likely did not significantly influence the results.

5.5. Relevance to the field & limitations of the results

Most previous studies on appearance in amblyopia lacked a control group, and did not adjust stimulus size for the visual acuity deficit in observers with amblyopia. In the present study, we accounted for acuity differences by scaling stimuli to each observer's (eye's) visual acuity, and found that despite these adjustments, observers with amblyopia showed reduced accuracy and structural similarity for crowded targets, and to some extent perceived stimuli as more distorted than controls. These findings, and the resulting target appearance dataset, provide a valuable resource for further investigations of amblyopia as well as other visual disorders. For example, for creating models to predict amblyopia from appearance responses (Desai et al., 2025), or for quantification of different error categories and qualitative analyses (Gomes Tomaz et al., in preparation).

We acknowledge that our sample was relatively small ($N = 15$ controls, $N = 14$ observers with amblyopia) and note that one of the observers with amblyopia was three letters short of our 0.2 logMAR interocular visual acuity difference but was included to help age-match the groups. Including larger numbers of participants will be important to determine whether the effects found here are dependent on the subtype of amblyopia. Inclusion of pseudo-letters or meaningless shapes in future studies will contribute to reducing possible effects of prior knowledge, such as biases to respond with letters. Future studies systematically varying target-flanker spacing while capturing target appearance will provide valuable insights into the mechanisms underlying crowding and appearance variability in amblyopia and other observer groups.

6. Conclusion

In summary, we employed a novel method to capture stimulus appearance in amblyopia and controls, investigating the perception of isolated and crowded stimuli. We controlled for visual acuity differences by adjusting stimulus sizes according to each individual's acuity threshold. Accuracy, structural similarity and perceived similarity were lower for observers with amblyopia compared to controls suggesting that observers with amblyopia perceive stimuli more distorted compared to controls even when the acuity deficit is taken into account. We also found significantly more crowding for observers with amblyopia with accuracy and structural similarity measures, but not with a perceived similarity index. This suggests that while appearance under crowding may be more distorted in amblyopia compared to controls, the reported shapes may still often be identifiable by human observers (raters) with free viewing and unlimited presentation times as a particular target letter, which could bias perceived similarity metrics and account for the absence of a Group x Crowding effect for perceived similarity. Finally, the relationship between accuracy, structural similarity, and perceived similarity underscores the value of including both performance and appearance based measures when investigating the limits of perception.

CRediT authorship contribution statement

Ângela Gomes Tomaz: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Wolf Harmening:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Dennis M. Levi:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Bilge Sayim:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by grants from the Agence Nationale de la Recherche (ANR-19-FRAL-0004 awarded to Bilge Sayim) and Deutsche Forschungsgemeinschaft (DFG project ID: 430279747, awarded to Wolf Harmening), as well as the National Eye Institute (R21EY030609 awarded to Dennis Levi) and a mobility grant from the University of Lille (MOBLILEX 2022-2023, awarded to Ângela Gomes Tomaz).

The experiment was not preregistered. Parts of this work were presented at the 45th European Conference on Visual Perception (2023), at the 24th annual meeting of the Vision Sciences Society (2024) and at the Bay Area Vision Research Day (BAVRD) 2024 conference.

We would like to acknowledge the undergraduate students – Jessica Lien, Minju Shim and Taylor Lui - who helped during a portion of the data collection, Adrien Chopin and Zeynep Yildirim-Keles for helpful comments on the analysis, Taylor Lui for her help rating over 17000 images, and Sunwoo Kwon for his feedback on posters previously presented on this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2026.108873.

References

Acar, K., Kiorpes, L., Movshon, J. A., & Smith, M. A. (2019). Altered functional interactions between neurons in primary visual cortex of macaque monkeys with experimental amblyopia. *Journal of Neurophysiology, 122*(6), 2243–2258. https://doi.org/10.1152/jn.00232.2019

Bach, M. (1996). The Freiburg visual acuity test - automatic measurement of visual acuity. *Optometry and Vision Science, 73*(1), 49–53. https://doi.org/10.1097/00006324-199601000-00008

Barrett, B. T., Pacey, I. E., Bradley, A., Thibos, L. N., & Morrill, P. (2003). Nonveridical visual perception in human amblyopia. *Investigative Opthalmology & Visual Science, 44*(4), 1555. https://doi.org/10.1167/iovs.02-0515

Birch, E. E. (2013). Amblyopia and binocular vision. *Progress in Retinal and Eye Research, 33*, 67–84. https://doi.org/10.1016/j.preteyeres.2012.11.001

Bonneh, Y. S., Sagi, D., & Polat, U. (2004). Local and non-local deficits in amblyopia: Acuity and spatial interactions. *Vision Research, 44*(27), 3099–3110. https://doi.org/10.1016/j.visres.2004.07.031

Bonneh, Y. S., Sagi, D., & Polat, U. (2007). Spatial and temporal crowding in amblyopia. *Vision Research, 47*(14), 1950–1962. https://doi.org/10.1016/j.visres.2007.02.015

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*(5241), 177–178. https://doi.org/10.1038/226177a0

Bradley, A., & Freeman, R. D. (1985). Is reduced Vernier acuity in amblyopia due to position, contrast or fixation deficits? *Vision Research, 25*(1), 55–66. https://doi.org/10.1016/0042-6989(85)90080-X

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Coates, D. R., Levi, D. M., Touch, P., & Sabesan, R. (2018). Foveal crowding resolved. *Scientific Reports, 8*(1), Article 9177. https://doi.org/10.1038/s41598-018-27480-4

Coates, D. R., Wagemans, J., & Sayim, B. (2017). Diagnosing the periphery: Using the Rey–Osterrieth complex figure drawing test to characterize peripheral visual function. *i-Perception, 8*(3), Article 2041669517705447. https://doi.org/10.1177/2041669517705447

Danilova, M. V., & Bondarko, V. M. (2007). Foveal contour interactions and crowding effects at the resolution limit of the visual system. *Journal of Vision, 7*(2), 25. https://doi.org/10.1167/7.2.25

Desai, K., Sayim, B., Levi, D. M., & Gomes Tomaz, Â. (2025). Predicting amblyopia and crowding from appearance captures. In *Vision Sciences Society Annual Meeting.* https://doi.org/10.1167/jov.25.9.1778. Florida, USA.

Farzin, F., & Norcia, A. M. (2011). Impaired visual decision-making in individuals with amblyopia. *Journal of Vision, 11*(14), 1–10. https://doi.org/10.1167/11.14.6

Flom, M. C., Weymouth, F. W., & Kahneman, D. (1963). Visual resolution and contour interaction. *Journal of the Optical Society of America, 53*(9), 1026. https://doi.org/10.1364/JOSA.53.001026

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Thousand Oaks CA: Sage https://www.john-fox.ca/Companion/.

Gabai, A., & Zeppieri, M. (2023). Anisometropia.. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK582146/.

Gomes Tomaz, Â., Daneshi, A., Hansmann-Roth, S., Harmening, W. M., & Sayim, B. (2022). The appearance of tiny objects: How Snellen's symbols look like below, at and above threshold. *Journal of Vision, 22*(14), 3696. https://doi.org/10.1167/jov.22.14.3696

Gomes Tomaz, A., Harmening, W.M., Levi, D.M., & Sayim, B. (in preparation). Qualitative and quantitative effects of crowding on visual appearance in amblyopia [Manuscript in preparation]. School of Optometry,. Berkeley: University of California.

Greenwood, J. A., Tailor, V. K., Sloper, J. J., Simmers, A. J., Bex, P. J., & Dakin, S. C. (2012). Visual acuity, crowding, and stereo-vision are linked in children with and without amblyopia. *Investigative Opthalmology & Visual Science, 53*(12), 7655. https://doi.org/10.1167/iovs.12-10313

Gunnlaugsdottir, E., Arnarsson, A., & Jonasson, F. (2008). Prevalence and causes of visual impairment and blindness in Icelanders aged 50 years and older: The Reykjavik eye study. *Acta Ophthalmologica, 86*(7), 778–785. https://doi.org/10.1111/j.1755-3768.2008.01191.x

Hariharan, S., Levi, D. M., & Klein, S. A. (2005). "Crowding" in normal and amblyopic vision assessed with Gaussian and Gabor c's. *Vision Research, 45*(5), 617–633. https://doi.org/10.1016/j.visres.2004.09.035

Hartig, F. (2024). DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression models. R package version 0.4.7 https://CRAN.R-project.org/package=DHARMa.

Herzog, M. H., & Sayim, B. (2022). Crowding: Recent advances and perspectives. *Journal of Vision, 22*(12), 15. https://doi.org/10.1167/jov.22.12.15

Herzog, M. H., Sayim, B., Manassi, M., & Chicherov, V. (2016). What crowds in crowding? *Journal of Vision, 16*(11), 25. https://doi.org/10.1167/16.11.25

Hess, R. F., Campbell, F. W., & Greenhalgh, T. (1978). On the nature of the neural abnormality in human amblyopia; neural aberrations and neural sensitivity loss. *Pflügers Archiv / European Journal of Physiology, 377*(3), 201–207. https://doi.org/10.1007/BF00584273

Hess, R. F., Dakin, S. C., Tewfik, M., & Brown, B. (2001). Contour interaction in amblyopia: Scale selection. *Vision Research, 41*(17), 2285–2296. https://doi.org/10.1016/S0042-6989(01)00099-2

Hess, R. F., & Jacobs, R. J. (1979). A preliminary report of acuity and contour interactions across the amblyope's visual field. *Vision Research, 19*(12), 1403–1408. https://doi.org/10.1016/0042-6989(79)90214-1

Kalpadakis-Smith, A. V., Tailor, V. K., Dahlmann-Noor, A. H., & Greenwood, J. A. (2022). Crowding changes appearance systematically in peripheral, amblyopic, and developing vision. *Journal of Vision, 22*(6), 3. https://doi.org/10.1167/jov.22.6.3

Kandel, G. L., Grattan, P. E., & Bedell, H. E. (1980). Are the dominant eyes of amblyopes normal? *Optometry and Vision Science, 57*(1), 1–6.

Kiorpes, L. (2019). Understanding the development of amblyopia using macaque monkey models. *Proceedings of the National Academy of Sciences, 116*(52), 26217–26223. https://doi.org/10.1073/pnas.1902285116

Kiorpes, L., & Daw, N. (2018). Cortical correlates of amblyopia. *Visual Neuroscience, 35*, E016. https://doi.org/10.1017/S0952523817000232

Kiorpes, L., & McKee, S. P. (1999). Neural mechanisms underlying amblyopia. *Current Opinion in Neurobiology, 9*(4), 480–486. https://doi.org/10.1016/s0959-4388(99)80072-5

Lenth, R. (2025). Emmeans: Estimated marginal means, aka least-squares means. R package version 1.10.6–090003 https://rvlenth.github.io/emmeans/.

Levi, D. M. (2008). Crowding—an essential bottleneck for object recognition: A mini-review. *Vision Research, 48*(5), 635–654. https://doi.org/10.1016/j.visres.2007.12.009

Levi, D. M. (2020). Rethinking amblyopia 2020. *Vision Research, 176*, 118–129. https://doi.org/10.1016/j.visres.2020.07.014

Levi, D. M. (2021). Amblyopia. In *, 178. Handbook of clinical neurology* (pp. 13–30). Elsevier. https://doi.org/10.1016/B978-0-12-821377-3.00002-7.

Levi, D. M., & Carney, T. (2011). The effect of flankers on three tasks in central, peripheral, and amblyopic vision. *Journal of Vision, 11*(1), 1–23. https://doi.org/10.1167/11.1.10

Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in amblyopic vision. *Vision Research, 42*(11), 1379–1394. https://doi.org/10.1016/S0042-6989(02)00061-5

Levi, D. M., & Harwerth, R. S. (1980). Contrast sensitivity in amblyopia due to stimulus deprivation. *British Journal of Ophthalmology, 64*(1), 15–20. https://doi.org/10.1136/bjo.64.1.15

Levi, D. M., & Klein, S. A. (1983). Spatial localization in normal and amblyopic vision. *Vision Research, 23*(10), 1005–1017. https://doi.org/10.1016/0042-6989(83)90011-1

Levi, D. M., & Klein, S. A. (1985). Vernier acuity, crowding and amblyopia. *Vision Research, 25*(7), 979–991. https://doi.org/10.1016/0042-6989(85)90208-1

Levi, D. M., & Klein, S. A. (2003). Noise provides some new signals about the spatial vision of amblyopes. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 23*(7), 2522–2526. https://doi.org/10.1523/JNEUROSCI.23-07-02522.2003

Levi, D. M., Klein, S. A., & Hariharan, S. (2002). Suppressive and facilitatory spatial interactions in foveal vision: Foveal crowding is simple contrast masking. *Journal of Vision, 2*(2), 140–166. https://doi.org/10.1167/2.2.2

Levi, D. M., Knill, D. C., & Bavelier, D. (2015). Stereopsis and amblyopia: A mini-review. *Vision Research, 114*, 17–30. https://doi.org/10.1016/j.visres.2015.01.002

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *The Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139

Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect Vernier thresholds. *Journal of Vision, 7*(2), 1. https://doi.org/10.1167/7.2.1

Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision, 12*(10), Article 13, 1–14 https://doi.org/10.1167/12.10.13.

Manassi, M., Sayim, B., & Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *Journal of Vision, 13*(13), Article 10, 1–10 https://doi.org/10.1167/13.13.10.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McGillycuddy, M., Popovic, G., Bolker, B. M., & Warton, D. I. (2025). Parsimoniously fitting large multivariate random effects in glmmTMB. *Journal of Statistical Software, 112*(1), 1–19. https://doi.org/10.18637/jss.v112.i01

Molaei, H., Farishta, R. A., & Farivar, R. (2025). Letter distortion mapping in amblyopia: Spatial patterns, stability, and relationship to visual acuity. *Investigative Ophthalmology & Visual Science, 66*(15), 65. https://doi.org/10.1167/iovs.66.15.65

Norgett, Y., & Siderov, J. (2017). Effect of stimulus configuration on crowding in strabismic amblyopia. *Journal of Vision, 17*(13), 5. https://doi.org/10.1167/17.13.5

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739–744. https://doi.org/10.1038/89532

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pelli, D. G., Levi, D. M., & Chung, S. T. (2004). Using visual noise to characterize amblyopic letter identification. *Journal of Vision, 4*(10), 904–920. https://doi.org/10.1167/4.10.6

Polat, U., Bonneh, Y., Ma-Naim, T., Belkin, M., & Sagi, D. (2005). Spatial interactions in amblyopia: Effects of stimulus parameters and amblyopia type. *Vision Research, 45*(11), 1471–1479. https://doi.org/10.1016/j.visres.2004.12.014

Popple, A. V., & Levi, D. M. (2008). The attentional blink in amblyopia. *Journal of Vision, 8*(13), 1–9. https://doi.org/10.1167/8.13.12

Pugh, M. (1958). Visual distortion in amblyopia. *British Journal of Ophthalmology, 42*(8), 449–460. https://doi.org/10.1136/bjo.42.8.449

Rappeport, N., Lev, M., & Polat, U. (2025). Foveal crowding increases with age under limited viewing conditions. *Scientific Reports, 15*, 42244. https://doi.org/10.1038/s41598-025-26451-w
Sayim, B., & Wagemans, J. (2017). Appearance changes and error characteristics in crowding revealed by drawings. *Journal of Vision, 17*(11), 1–16. https://doi.org/10.1167/17.11.8
Sayim, B., Westheimer, G., & Herzog, M. H. (2008). Contrast polarity, chromaticity, and stereoscopic depth modulate contextual interactions in Vernier acuity. *Journal of Vision, 8*(8), 1–9. https://doi.org/10.1167/8.8.12
Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science, 21*(5), 641–644. https://doi.org/10.1177/0956797610368811
Schor, C., & Hallmark, W. (1978). Slow control of eye position in strabismic amblyopia. *Investigative Ophthalmology and Visual Science, 17*(6), 577–581.
Scialfa, C. T., Cordazzo, S., Bubric, K., & Lyon, J. (2013). Aging and visual crowding. *The Journals of Gerontology: Series B, 68*(4), 522–528. https://doi.org/10.1093/geronb/gbs086
Shamsi, F., Liu, R., & Kwon, M. (2022). Foveal crowding appears to be robust to normal aging and glaucoma unlike parafoveal and peripheral crowding. *Journal of Vision, 22*(8), Article 10, 1–13 https://doi.org/10.1167/jov.22.8.10.
Sharma, V., Levi, D. M., & Coletta, N. J. (1999). Sparse-sampling of gratings in the visual cortex of strabismic amblyopes. *Vision Research, 39*(21), 3526–3536. https://doi.org/10.1016/S0042-6989(99)00028-0
Sharma, V., Levi, D. M., & Klein, S. A. (2000). Undercounting features and missing features: Evidence for a high-level deficit in strabismic amblyopia. *Nature Neuroscience, 3*(5), 496–501. https://doi.org/10.1038/74872
Simmers, A. J., Gray, L. S., McGraw, P. V., & Winn, B. (1999). Contour interaction for high and low contrast optotypes in normal and amblyopic observers. *Ophthalmic and Physiological Optics, 19*(3), 253–260. https://doi.org/10.1046/j.1475-1313.1999.00416.x
Sireteanu, R., Lagreze, W.-D., & Constantinescu, D. H. (1993). Distortions in two-dimensional visual space perception in strabismic observers. *Vision Research, 33*(5–6), 677–690. https://doi.org/10.1016/0042-6989(93)90188-3
Song, S., Levi, D. M., & Pelli, D. G. (2014). A double dissociation of the acuity and crowding limits to letter identification, and the promise of improved visual screening. *Journal of Vision, 14*(5), 1–37. https://doi.org/10.1167/14.5.3
Stuart, J. A., & Burian, H. M. (1962). A study of separation difficulty: Its relationship to visual acuity in normal and amblyopic eyes. *American Journal of Ophthalmology, 53*(3), 471–477. https://doi.org/10.1016/0002-9394(62)94878-X
The MathWorks Inc.. (2022). *MATLAB version R2022a*. Natick, Massachusetts: The MathWorks Inc.. https://www.mathworks.com
Tsirlin, I., Colpa, L., Goltz, H. C., & Wong, A. M. F. (2018). Visual search deficits in amblyopia. *Journal of Vision, 18*(4), 17. https://doi.org/10.1167/18.4.17
Verghese, P., McKee, S. P., & Levi, D. M. (2019). Attention deficits in amblyopia. *Current Opinion in Psychology, 29*, 199–204. https://doi.org/10.1016/j.copsyc.2019.03.011
Von Noorden, G. K. (1981). New clinical aspects of stimulus deprivation amblyopia. *American Journal of Ophthalmology, 92*(3), 416–421. https://doi.org/10.1016/0002-9394(81)90534-1
Wang, Y., Zhang, B., Tao, X., Wensveen, J. M., Smith, E. L., & Chino, Y. M. (2017). Noisy spiking in visual area V2 of amblyopic monkeys. *The Journal of Neuroscience, 37*(4), 922–935. https://doi.org/10.1523/JNEUROSCI.3178-16.2016
Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing, 13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861
Waugh, S. J., & Fronius, M. (2024). Landolt C-tests with “fixed” Arcmin separations detect amblyopia but underestimate crowding in moderate-to-severe amblyopic children and adults. *Investigative Ophthalmology & Visual Science, 65*(10), 33. https://doi.org/10.1167/iovs.65.10.33
Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15*(4), 160–168. https://doi.org/10.1016/j.tics.2011.02.005
Wiesel, T. N., & Hubel, D. H. (1963). Effects of visual deprivation on morphology and physiology of cells in the cat’s lateral geniculate body. *Journal of Neurophysiology, 26*(6), 978–993.
Wiesel, T. N., & Hubel, D. H. (1965). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *Journal of Neurophysiology, 28*(6), 1029–1040. https://doi.org/10.1152/jn.1965.28.6.1029
Williams, D. R. (1985). Aliasing in human foveal vision. *Vision Research, 25*(2), 195–205. https://doi.org/10.1016/0042-6989(85)90113-0
Wong, E. H., Levi, D. M., & McGraw, P. V. (2005). Spatial interactions reveal inhibitory cortical networks in human amblyopia. *Vision Research, 45*(21), 2810–2819. https://doi.org/10.1016/j.visres.2005.06.008
Worth, C. A. (1903). *Squint: Its causes, pathology and treatment*. Philadelphia: The Blakiston Co.
Wright, K. W. (2006). Visual development and amblyopia. In K. W. Wright, P. H. Spiegel, & L. S. Thompson (Eds.), *Handbook of pediatric strabismus and amblyopia* (pp. 103–137). Springer. https://doi.org/10.3368/aoj.56.1.200-a.
Zele, A. J., Pokorny, J., Lee, D. Y., & Ireland, D. (2007). Anisometropic amblyopia: Spatial contrast sensitivity deficits in inferred magnocellular and parvocellular vision. *Investigative Opthalmology & Visual Science, 48*(8), 3622. https://doi.org/10.1167/iovs.06-1207