Amblyopia: assessment and treatment of binocular visual function

Manuela Bossi

Thesis submitted to the University College of London
for the Degree of *Doctor of Philosophy*
- Vision Neuroscience -

March, 2018

Supervisors:
Professor Steven C. Dakin
Professor Gary Rubin

Funding:
Supported by UCL Impact Awards & Special Trustees of Moorfields Eye Hospital, London.

(SN: 13019710)

I, Manuela Bossi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

To my dad

(… and our “magical mistery tour” …)

Acknowledgments:

- ✓ Background, Methods, Results, Discussion and Conclusion - to reach the expected and desired structure.
- ✓ Passion, commitment, precision and (effortful) dedication - to insert all the relevant content.
- ✓ Logic, creativity, criticism and style - to give authority and credibility.

Furthermore, beyond personal control,

- ✓ support, inspiration and co-operation - to obtain great results.

In my opinion, I have just listed most of the main factors contributing to a thesis.
This thesis is the conclusion of a once-in-a-life experience.
Commonly, we think of life as a book to be written... [Not good enough] ...
Let's think of life as a *thesis* to be written.

First, I would like to immensely thank my family, who gave me all the background needed to contextualise a true life-experience…to be lived in full (or 'al massimo').
Then, thanks to all my friends, 'recruited' from anywhere, based on one broad inclusion-criterion: be willing to grow together, going through multivariate experiences.
And thanks to the research team, in particular to my supervisor and a few colleagues, who significantly helped to demonstrate the results of this journey.
Thanks to the one who shared with me everything in my last (~10) years: we have collected and discussed so many data together… Further research is welcome (and/or needed) ☺

In particular, special thanks go to:

(my family)
MammaMi, unmissable; fratellino, my 'other half'; nonni Lino&Anna, my half-family; and a special hug to Hegel, the only who has always *silently* supported me. Thanks to all the rest of my family too, who will never stop reminding me how difficult it is to explain 'what kind of work do you exactly do?'

(my friends)
Riccardo, Federico e Pilli, who never missed to be there for me, in their own ways; Alessandro, Alessandra, Stefania, Andrew, Riccio, Ale and Silvia, who were always ready for a drink/tea at the right time (despite the difficulties in organising it, sometimes); Bilal, Michele, Charu, Nilpa, Cristina, Mika, Mortiz and Eylem, who reminded me the importance of what I was doing by showing me the passion for their work; James, for being himself; Jon, who never stopped believing in me; Chris, my official 'English' problem-solver; Rob, Rob, Gordon, Adam and Vicki for making my stay in London as troubled (sometimes) as special. And thanks to all the ones I haven't mentioned but still deserve a 'thank you'.

(the research team)
Annegret, Gary, John and Vijay, plus Susie and Nick, who all collaborated at different stages and in different ways to my project; Elaine, Nilpa, Phil, Bilal and Andy, who were fundamental in certain moments, & Lisa and Moritz, who definitely were - or I would have never get my last experiment done; and thanks to Steven, who made this whole journey possible, real and unique, with the ups and downs we should all expect (and be ready to handle) in life.

(my partner)
Gabriele, and the multiple roles he had, has & hopefully will have in my life.

(…)
Now, my words would need to be too clear and too loud to reach you, far away from anywhere I would want them to arrive…still, you're so close to me that there is no need to say a word. Love my family, love you, dad. We did it. Thanks for everything.

… with the unaccountable added value that my PhD experience (only partially wrapped up in this thesis) has given to my knowledge, my skills and my 'scientific life', I am looking forward to a new chapter of my magical mystery tour.

Thank you all!

Abstract

Short ~

Unilateral amblyopia is a common neurodevelopmental syndrome characterized by reduced acuity and contrast sensitivity in the amblyopic eye (AE) & by abnormal inter-ocular visual function, e.g. reduced stereoacuity; without a concomitant etiological dysfunction. Standard treatment consists of a period of optical correction followed, when necessary, by occlusion therapy. Although ~70% children gain vision, this monocular therapy is limited by poor compliance and uncertain impact on stereo-function. Recently, binocular treatments have attempted to “rebalance” vision, by adjusting the intensity of monocular visual inputs (enhancing usage to AE or reducing fellow-eye -FE- one), while stimulating binocular cortical interactions.

We have developed a “Balanced Binocular Viewing” (BBV) treatment that has patients spend an hour per day at home watching modified movies while wearing 3D goggles (to control what each eye sees). Movies present a blurred image to the FE and a sharp image to the AE. Performance (compliance and binocular-imbalance) is monitored throughout treatment using the child’s performance on a game, played during movie playback. Two ‘ghost’-stimuli, each made of a mixture of luminance increment/decrement, were presented dichoptically (some visible only through goggles): we quantified the mixture required for the child to be equally likely to report either ghost as ‘whiter’. Treating children (N=22) for 8-24 weeks lead to significant improvement in the AE acuity (mean gain: 0.27 logMAR). This is comparable to results achieved with occlusion, but elicits much higher compliance (89% of prescribed daily dose).

We also compared our measure of binocular-imbalance to others, also quantifying sensory eye-dominance, to assess any test’s suitability to complement clinical practice. Pilot data measured with adult and children, with and without amblyopia, suggest that a variant of the ’ghost’-game is a potentially useful and efficient stand-alone clinical test with the advantage of being suitable for unsupervised home-based monitoring of patient’s binocular status.

Long ~

Amblyopia is a common developmental disorder of spatial vision caused by a prolonged imbalance in visual stimulation in early childhood. For example, a difference in blur across the eyes (in *anisometropic amblyopia*, i.e. reduced refractive power in the amblyopic/affected eye) and/or interocular misalignment (in *strabismic amblyopia*) leads to interocular differences (IOD) in visual function. The differences persist following optical correction despite the absence of any ocular or neurological pathology. The mechanism of amblyopia is poorly understood. Possibly, the reduced neuronal response to stimuli presented to the affected eye (AE) is a consequence of either a loss or a reduced excitability of neurons in primary visual cortex (V1). The functional deficit associated with the amblyopic syndrome includes reduced visual acuity and contrast sensitivity in the AE, vulnerability to foveal crowding in the AE and poor stereoacuity. Left untreated amblyopia can have negative impact on quality of life. Current clinical best-practice is a period of refractive adaptation (resolving the deficit in at least 28% of 3-7 yrs children, after 18 weeks - Cotter, Foster et al., 2012; mean duration ~15w.- Stewart, Moseley, Fielder et al., 2004) followed by occlusion of the less affected fellow eye (FE), through patching or pharmacological intervention. Such treatments enjoy good success, as ~70% children gain at least ~0.2 logMAR (Stewart, Moseley, & Fielder, 2011). But they also have several limitations: poor compliance (>50% fail to comply; M. P. Wallace, Stewart et al., 2013), high rate of recurrence (≥0.2logMAR loss in ~30% at1year; Bhola, Keech et al., 2006) and unlikely impact on stereo-function (<50%; Stewart, Wallace et al., 2013). Recently, binocular treatments have attempted to “rebalance” asymmetrical visual input based on the notion that a central suppression of AE input (beyond V1) might be responsible for both monocular and binocular functional deficits in amblyopia. Binocular treatments work by adjusting the intensity of monocular visual inputs (either increasing AE stimulation or reducing FE one) to promote normal binocular interactions (reweighting of neural connections in binocular cortex).

This project describes a new *“Balanced Binocular Viewing” (BBV)* treatment for amblyopia. BBV requires patients to spend an hour per day at home watching movies while wearing modified 3D shutter-glasses. The glasses allow us to present dichoptic visual stimuli where the FE receives an image Gaussian blurred to a level to elicit crowded acuity comparable to the AE. Images are slowly modulated in horizontal disparity in order to promote stereovision. Additionally, children’s performance on a simple video game (played periodically during movie playback) allow us to monitor both compliance (is the child wearing the goggles?) and the level of binocular-imbalance (how much does the child favour one eye?).

Our exploratory study on children (24 children aged 6 to 11 yrs; with anisometropic and/or strabismic amblyopia) showed that a period of 8 to 24 weeks of treatment (following, on average, 28 weeks (σ:12) of refractive adaptation) leads to a significant improvement in visual acuity (N=22; 0.27 logMAR gain in the AE, on average; Keeler chart) and elicits good compliance (on average, 89% of prescribed daily dose on 68% of days when the system was installed at home). Contrary to the notion that intra-ocular suppression (IOS) may have a causative role in amblyopia, IOS did not change significantly following treatment.

A key component of our home-based therapy was its inclusion of a *psychophysical method for quantifying binocular balance* (our index of IOS). We next sought to compare our measure to other measures and specifically to assess their suitability for development as a clinical *test of binocular balance*. When comparing the same tests on adults with normal vision we report that a variant of the test we used for monitoring IOS - involving a brightness judgement of superimposed opposite-contrast polarity same-identity optotypes - exhibits superior test reliability (quantified using intra-class correlation) than comparable tests. Pilot data measured with both adult and children with amblyopia suggest that this test, along with the letter tests derived from Kwon, Wiecek et al. (2015), are potentially useful.

In summary, we have described a new home-based binocular therapy that engages high levels of compliance in children and leads to gains in acuity comparable to those achieved with equal or longer periods of patching. We have also gone on to show that the test of binocular balance developed for monitoring IOS during treatment is a candidate for being an efficient stand-alone clinical test that has the advantage of being suitable for unsupervised home-based monitoring of children's binocular status.

Contents:

List of Figures:

List of Tables:

1 Background

1.1 Structure of the human visual system

From the retina

The wavelength of visible-light falls in the range 400 to 700 nm. The retina contains two classes of photoreceptors that transduce light into electrical signals: cones – specialized in spatial resolution and sensitive to photopic light levels ($10\text{-}10^8$ cd/m^2) - and rods, specialised for sensitivity to scotopic ($10^{-3}\text{-}10^{-8}$ cd/m^2) light levels (Purves, Augustine, & Fitzpatrick, 2001). Within the retina, photoreceptors are connected to bipolar and amacrine cells which then project to ganglion cells. The area of visual space where light modulates neural activity is known as the receptive field (**RF**) of the cell and its extent depends on the intensity, frequency and size of the spot of light received (Hartline, 1938). Bipolar (and ganglion) cells have a centre-surround structure of RFs ("centre-ON" or "centre-OFF") where e.g. light falling in the RF's centre will excite the neuron (thus, "centre-ON") but light falling in the surround will inhibit activity (Dacey, Packer et al., 2000). The arrangement of inputs determines if a sub-region of the cell's RF leads to an inhibitory or excitatory drive. In retinal ganglion cells: a depolarization of the pre-synaptic (e.g. bipolar) cell induces a positive post synaptic potential that leads to an increase in the retinal ganglion cell's firing, while a hyperpolarization of the pre-synaptic cell inhibits post synaptic potential (Kuffler, 1953).

Ganglion cells are broadly classed as parvocellular (*P-cells*, covering the fovea and parafovea, with smaller cell bodies), magnocellular (*M-cells*, distributed more densely throughout the peripheral retina with larger cell bodies), or koniocellular (K-cells, or nonM-nonP, with very large non-concentric RF) - see Casagrande (1994) for a review.

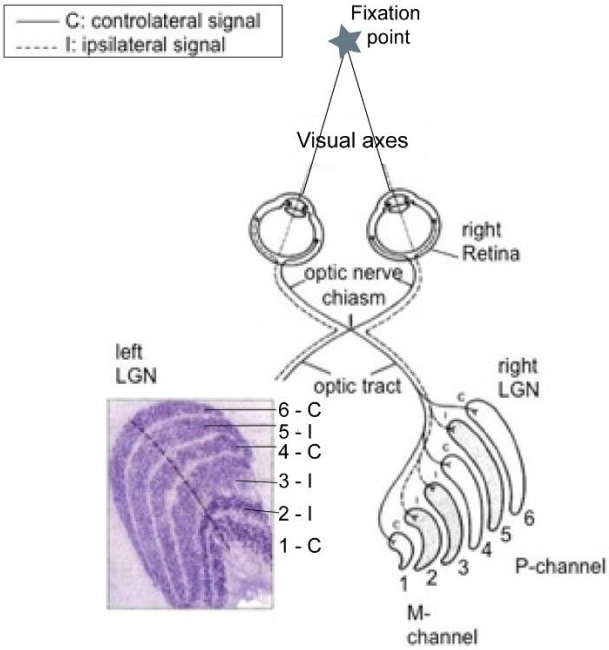


Figure 1 From the retina to LGN - **From the top** of the image: light enters each eye and stimulates the retina of each eye. Through the optic nerve, temporal and nasal retinal projections from each eye pass the 'chiasm' (and only retinal projection decussate on the contralateral circuit) and reach LGN (lateral geniculate nucleus) of the ipsilateral thalamus. LGN is a six-layers structure receiving input from M-cells (layers 1,2) and P-cells (layers 3 to 6) from the contralateral eye, by nasal retinal projections ('c'; layers 1,4,6), or ipsilateral eye, by temporal retinal projections ('i'; layers 2,3,5, shadowed). **Bottom left**: a coronal section of the left LGN of a cat. **Bottom right**: schematic reconstruction of the right LGN, with blank and shaded layers indicating selectivity for projections from the contralateral (1, 4, 6) or ipsilateral (2,3,5) eye respectively. Adapted from the drawings by Santiago Ramon y Cajal. From *Histologie du Systeme Nerveux de l'Homme et des Vertebres*, Madrid, 1952 – considered no longer under copyright protection (https://archive.org/details/histologiedusyst01ram).

To the Lateral Geniculate Nucleus (LGN)

About 90% of connections from the retina project to the lateral geniculate nucleus (LGN) within the thalamus[1]. The optic nerve terminates in the chiasm and (via a 'partial decussation' to the optic tract) connects to the ipsilateral LGN (only nasal-retinal connections cross-over). The visual pathway then continues through the optic tract to the primary visual cortex, conveying signals from the contralateral hemi-field (see Figure 1). The LGN has six layers where right/left eye and M/P/Konio signals remain separated, as shown in Figure 1. Layers 1, 2 serve the M-channel conveying a rapid-transient response; upper layers (3, 4, 5, 6) process P-signals, with slower and sustained response; a thin substrate under each layer drives Konio information. In LGN, strati 1-

[1] The remaining (around 10%) of retinal connections, project to the superior colliculus and are involved in the control of eye movements. The pupillary reflex, as well, is fast-way controlled directly via retina-hypothalamus and –pretectum, regulating circadian rhythms and sleep quality; Bear, Connors, and Paradiso (2001)

4-6 receive contralateral temporal signal, while 2-3-5 layers the ipsilateral nasal signal (Bear et al., 2001). Like retinal ganglion cells, LGN cells have centre-surround RFs confers selectivity for *spatial frequency*-**SF**, the scale of detail in an image (Bear et al., 2001) – see also Figure 10.

To primary visual cortex

From LGN, visual signals project into layer 4 of primary visual cortex (V1 or striate cortex; Figure 2), subdivided into 4A, 4B, 4C (4Cα and 4Cβ). V1 has six layers and is *retinotopically* organized (see 1.2, "Ice-Cube" model). V1 cells, by contrast to retina and LGN, have larger more elongated RFs (see Figure 4).

Three **parallel visual pathways** can be distinguished (summarised in Yoonessi & Yoonessi, 2011):

1) "M channel": M-ganglion cells, through layers 1, 2 in LGN, reach 4Cα in V1, where monocularity is preserved. Layer 4Cα is connected to 4B where binocular information is integrated and contrast-properties extracted. This channel is primarily concerned with encoding visual motion.

2) "P channel": P-ganglion cells, through layers 3, 4, 5, 6 in LGN, reach 4Cβ in V1; **binocular** information is integrated in 4A and conveyed to strati 2, 3 in interblob-columns, where "complex cells" process phase-properties. This channel is concerned with encoding the *shape* of visual stimuli.

3) "Blob channel": K-retinal cells project to 6 LGN thin sub-layers and then into V1 so-called blob columns, where cells with circular RFs process colour. Note the P-channel occupies the "inter-blob" regions.

M & P *channels* are also classed as *transient* and *sustained*, respectively, since they respond differently to temporal modulation of grating-stimuli (transient: flickering; sustained: static) (Hubel & Wiesel, 1968). These pathways are thus thought to support sensitivity to high spatial frequencies – SFs (P; transient) and low SFs (M; sustained), as introduced by Kulikowski and Tolhurst (1973).

Cell's property	**M-ganglions**	**P-ganglions**	**K-ganglions**
RF's structure	on-off centre/surround	on-off centre/surround	on-off centre/surround
cell's body	large	small	small
speed of conduction	fast	slow	very slow
SF sensitivity	low SF	high SF	mid-high SF
light sensitivity	high (photopic)	low (scotopic)	mid-high
channel	**Parvocellular**	**Magnocellular**	**Koniocellular**
LGN connection layers	1,2	3,4,5,6	sub-layers (1to6)
V1 connection layers	IVCα-IVB-II/III(blobs) – V/VI	IVCβ-IVA-III/II (interblob)	blobs (II-III)
Extrastriate areas	V2-V3-MT-MST-(LIP-VIP-STP-7a)	V2-V3-V4-IT	
cortical stream	dorsal	ventral	(ventral)

Table 1 Summary of the main properties (left column) of three types of retinal ganglion cells along with their downstream channel-properties.

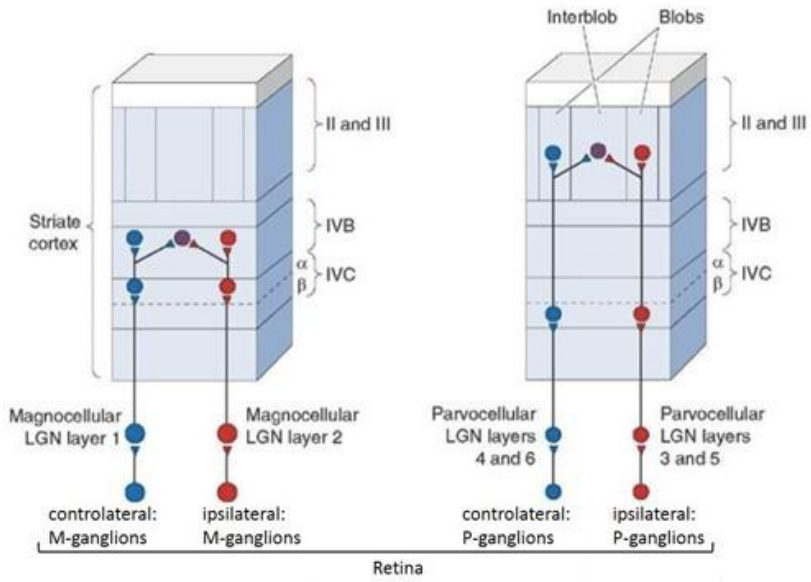


Figure 2 Schematic reproduction of retino-cortical projections. **On the left**: M-channel. From the bottom, neural signal from M-ganglion cells is conveyed to layers 1, 2 of the LGN, maintaining separate contralateral (blue) and ipsilateral (red) signals. The projection continues to layer 4Cα of V1 where binocular connections are established in upper layer 4B. **On the right**: P-channel. From the bottom, neural signals from P-ganglion cells originating in the controlateral retina are conveyed to LGN layers 4, 6 while P-ganglions of the ipsilateral retina projects to LGN layers 3, 5. In cortex, V1 is reached in its layer 4Cα and then binocular connections are established into inter-blob columns of upper layers 2 and 3. K-channel is not represented, because it projects from P-cells into all six sub-layers in LGN; then reaches blobs (upper right-figure) in layers 2 and 3 of V1. Adapted from http://what-when-how.com/, by Crankshaft's staff - using articles which are in contract with several publishing houses, on revenue share basis.

Beyond V1

From V1, the visual information 'splits' into a "two-stream pathway": a *dorsal stream* leading to parietal areas, and a *ventral stream* towards temporal areas *(*Figure 3*)*. A hierarchical organization of visual processing has been proposed: ventral for *object recognition*, and dorsal for *spatial localisation and motion* (Bear et al., 2001). Based on studies on rhesus monkeys, the ventral pathway have been said to perform *perception*

(involving as it does V4, which encodes colour, shape and orientation, and inferotemporal area, encoding faces and objects) and the dorsal pathway-*action* (involving as it does V5/MT which encodes movement) (Goodale & Milner, 1992). A strict dichotomy of two-pathways has been superseded by a better understanding of *modular* and *hierarchical* organisation of the visual system (i.e. functional and cyto-architectural sets of operation computed in progressive downstream areas).

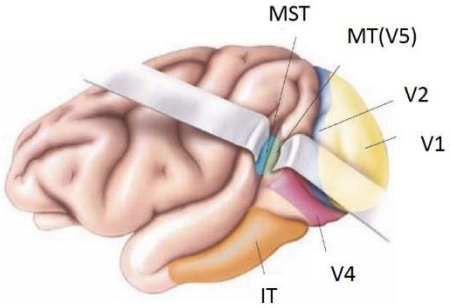


Figure 3 Cortical visual areas active in the human visual system (left hemisphere view). Under the "two-pathway" view, the dorsal stream proceeds from V1 to V2, to MT (medio temporal parietal, corresponding to area V5 in monkeys) and on to MST (superior medio temporal area). The ventral stream passes from V1 to V2 to extra-striate areas, V4 and IT (inferotemporal cortex). In the case of the dorsal stream, circuits are thought to elaborate mainly motion information for action programming while in the case of the ventral stream, they are involved in perception for object recognition. Adapted from http://what-when-how.com/, by Crankshaft's staff– original source: Mishkin, Ungerleider, and Macko (1983), p.414 – licence agreement # 4414800673684.

1.2 Functional architecture of V1

During the 1950's and 60's, D.H. Hubel and T. N. Wiesel expanded our knowledge of RF organisation and V1 functional architecture (from physiological studies in kittens). They classed V1 cells as *simple cells* (monocularly driven; predominantly located in V1-layer 4 and deep layer 3) and *complex* cells (binocularly driven; higher orders and distributed through upper and lower layers, 5-6 and 2-3) - Hubel and Wiesel (1962), see Figure 2 and Figure 4. Both classes of cells, in contrast to retinal and LGN cells have elongated RFs that are orientation selective. However only complex cells show *spatial invariance*, i.e. they are not selective for the contrast-polarity of features falling within their RFs. Simple cells respond to stimuli falling in specific locations within excitatory/inhibitory portions of the RF, and sensitive to the phase of their inputs. The output of simple cells is thought to be aggregated by complex cells, which respond to features falling in any position within the RF. For a review, see Hubel (1982). In terms of the dynamics of their responses, simple cells show transient activity and complex

cells show sustained activity, i.e. complex cells are phase-insensitive (De Valois, Yund, & Hepler, 1982).

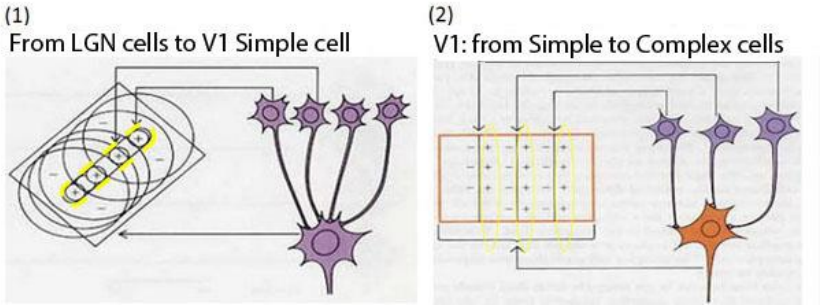


Figure 4 Schematic representations of receptive field (RF) organisation within V1, adapted from Hubel and Wiesel (1962), p. 142-143 - copyright licence agreement No 4414811341577. **(1)** Four LGN cells have centre-surround RFs (here "on-centre", optimally excited by a light disk). This arrangement collectively stimulates a V1 simple cell (in purple) that consequently responds to a specific orientation (marked by the yellow dashed line). **(2)** Simple cells (three, in purple) have RFs subdivided in excitatory (+) and inhibitory (-) regions. A complex cell (in orange) will be excited by any vertical stimulus falling across the area corresponding to multiple simple cells' RFs (e.g. the outlined orange box), regardless of its position or contrast polarity.

The same authors described V1 functional architecture as modular. Each module contains information about all possible orientations of elements from both eyes. In V1, 85% of cells are binocular, responding well to stimulation from either eye, although there is variation in the degree of eye dominance. The variation is sharp in layer 4 and progressively reduces in upper layers (Hubel & Wiesel, 1962). Cells with a preference (of any degree) for a given eye are organised into ocular dominance (**OD**) columns (Hubel, 1982). The "Ice-cube" model (Figure 5) proposes that a 'cube' of visual cortex (2*2mm in layer 4, with dimensions scaling with RFs size in other layers) fully elaborates a specific portion of visual space (~1deg^2), completely encoding eye preference (Figure 5: 'OD' arrow) and orientation (Figure 5: 'Orient.' arrow). Different units having the same monocular preference, i.e. falling in the same OD column, represent all the possible orientations. In addition, cells (mainly, complex cells) tuned for the same preference of motion-direction are clustered in columnar arrays: in some regions the cells respond equally well to the two opposite directions of movement, but in others there is a mixture of cells favouring one or the other direction (De Valois et al., 1982). Hubel and Wiesel found that across the visual field of rhesus monkeys, a light stimulus falling 20 degrees from the fovea (*centralis*) activated RFs so that each visual degree is represented by ~6mm of cortex, whereas only ~0.15 mm of cortex represents the same area of the peripheral visual field, until 80-90 degrees (Hubel & Wiesel, 1968). This "***cortical magnification***" defines the distance in cortex corresponding to 1 degree in the visual field so that the cortical representation is scaled to RFs properties (i.e. fovea has the highest resolution and magnification decreases for peripheral stimuli). Instead, *shape*

and orientation of the visual stimulus (in any position of the visual field) are uniformly represented in the cortex.

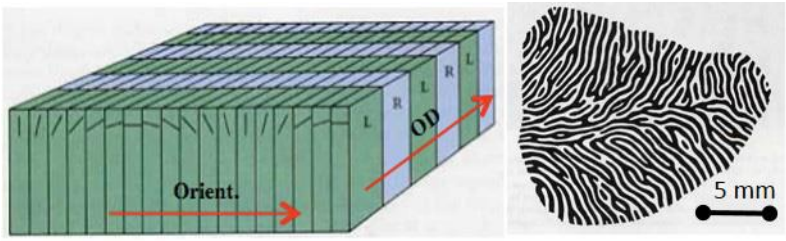


Figure 5 Left: The "Ice-cube" model (originally introduced by Hubel, 1982): schematic representation of the functional organisation of V1. Cells in V1 are organised into columns sharing common (1) Ocular Dominance (OD) and (2) orientation preference. Cells within (1) OD columns (vertically oriented), preferentially respond to right- or left-eye stimulation (in the figure: L green, R blue columns), roughly covering one cycle every mm. (2) Orientation units (arranged parallel to the surface) preferentially fire for each possible orientation, with preferences distributed on average in 10 deg steps across each OD column, e.g. ~10° shift in orientation-preference, clockwise or counter-clockwise, each 0.05 mm of cortex. **Right:** Schematic reconstruction of a tangential section of right V1 of a macaque, layer4 (C and B), originally made using the reduced-silver-stain method by LeVay, Hubel, and Wiesel (1975). Here, black and white stripes represent the alternating OD columns (right or left eye-preference).

Therefore, our visual system maps in cortex a 2D topographic representation (integrating information from both eyes) of the corresponding 3D visual space (approximately up to ±80 degrees of visual angle from fixation).

1.3 Binocular vision

The external visible space is referred to as visual field (VF) of the eyes. Each eye's VF overlaps with the other in the binocular central field (with right eye's field extending further to the right, and vice versa). Features in each VF stimulate specific portions of the retina. When features land in the binocular VF, they stimulate *corresponding* retinal points on each eye. Importantly, this topographical mapping of visual space (known as the *retinotopic map*) is maintained in the visual cortex.

Binocular summation *(or convergence)* is the cortical mechanism through which monocular signals are integrated across the eyes into a single percept. In fact, via a combined activity of eyes' muscles, each eye simultaneously turns either inward (*convergence*, for close objects) or outward (*divergence*, for further objects) to direct both visual axes towards the fovea. Concomitant accommodation of the lens regulates the refraction of the light-stimulus. Together, these mechanisms allow us to *focus* on a fixation target and so, by combined activity of ocular muscles and neurons, the slightly

different images from the two eyes are merged into a single percept. This process is referred to as (binocular) fusion.

With visual axes aligned, the target image stimulates both foveae in *correspondent retinal points*. Before V1 neural activity at these locations is monocularly segregated but within V1, specifically from layers 4β (M-channel) and 4α, 2 and 3 (P-channel), signals from each eye converge into RFs of binocular cells (primarily, complex-cells in V1). These cells are sensitive to retinal image disparity (see next paragraph) and activation of cells within the binocular visual field support precise representation of depth information. When a stimulus activates *non-correspondent* retinal points (i.e. not at equal distance from each eye's fovea), the observer could, in theory, experience *diplopia* (or double vision; see also 1.3.1) - although in reality we are equipped for ignoring/filtering this information (e.g. reallocating attention somewhere else). However, prolonged adaptation to anomalous correspondence (e.g. due to misalignment of the eyes) can lead to a neural compensation for the angle of deviation leading to fusion supported by *abnormal retinal correspondence*. Under these conditions input from one eye might either be *intermittently* preferred over input from the other eye (fixation preference), or there may be a *constant* preference for the output of one eye, inducing functional suppression of the non-dominant eye (see 1.3.1).

Depth/disparity tuning

Vergence movements determine the extent of deviation of the visual axis from the physical projection of the fixation point, influencing the horizontal disparity of monocular images falling in correspondent retinal points. This disparity supports stereopsis (Freeman & Ohzawa, 1990), commonly defined as the ability to judge depth based on binocular information (retinal disparity).

Stereopsis is sometimes defined as coarse- or fine-stereopsis (large or small disparities processed in the respective RFs), the last being clinically measured as *stereoacuity* (Poggio & Poggio, 1984). Disparities are classed as *crossed* disparity (disparity<0) arising from objects presented closer to the observer than the focal point (disparity=0) and *uncrossed* disparity (disparity>0) for further objects, giving a precise indication (retinal disparity) for object's position in the visual field (see Figure 7A).

The portion of the visual field that yields single vision, allowing for retinal correspondence and fusion is the *horopter.* G. Vieth (1818) and J. Muller (1823) argued that the horopter falls on a "circle" where all the points have the same angles when

individually projected into the two eyes (Vojnikovic & Tamajo, 2013). There is a zone around the horopter known as Panum's area, within which **fusion** of binocular information is possible, allowing for singleness of binocular vision (Ogle K., 1950 *Researches in Binocular Vision;* Philadelphia: Saunders). Particularly, we distinguish motor fusion (eye movements that allow convergence) and sensory fusion (neural activity of binocular cells).

An early study in cats established that the horizontal disparity eliciting maximum cortical response in correspondent points ranges 6.6 degrees of visual angle; less extended is instead the vertical disparity range, equal to 2.2 degrees (Barlow, Blakemore, & Pettigrew, 1967). However, binocular fusion would be possible within a *disparity gradient* (depending on the proximity of objects in the visual field) and not within an absolute extension of the fusional area (Braddick, 1979). This *gradient* has been defined as "a difference between the disparities of neighbouring objects divided by their angular separation", so that fusion could fail even at very small disparities when there are more objects near one another (Burt & Julesz, 1980).

An archetypal *binocular disparity* detector[2] (Ohzawa, Deangelis, & Freeman, 1990), reproduced in Figure 6, should: (i) receive both monocular inputs from correspondent retinal points, (ii) be selective for all and only the stimulus' position falling within its RF and (iii) be phase-dependent. These last two points respectively relate to the fact that each eye perceives a different perspective of the same object and the same binocularly viewed object maintains the same polarity. In cats, such a response arises in a subset of complex-cells, thought to be optimal detectors for binocular-disparity, with a specific selectivity for SF (i.e. size) acquired from their simple-cell input (Anzai, Ohzawa, & Freeman, 1999). Other studies showed that disparity-tuned cells are present in the visual cortex of humans (Ohzawa et al., 1990) and monkeys (Hubel & Wiesel, 1970a; Poggio & Poggio, 1984). Hubel and Wiesel (1970a) proposed the existence of cells tuned for *near* and *far* disparities (selectively activated by objects nearer or further away from the focal point) or at different relative depths. Poggio *et al* used dichoptically presented bars (of optimal orientation and size) to distinguish *near (crossed)* and *far*

[2] To investigate the mechanism of binocular summation, optimal stimuli for isolated cells (in terms of their orientation, SF, temporal frequency-TF) are identified under monocular presentation. Then, to identify a binocular-response, the same cells' sensitivity is measured (response amplitude; spike/sec) using optimal stimuli presented dichoptically (i.e. each eye is simultaneously presented with a modulated version of the same stimulus, e.g. drifting gratings at 2Hz, phase-shifted between eyes).

(uncrossed) disparity-tuned cells in V1 and V2 (predominantly simple and complex cells, respectively). About half of them were *tuned*, showing largely excitatory responses, for stimuli presented within the range of preferred disparity. Cormack and colleagues used a psychophysical task to probe disparity-channels, presenting sub-threshold stimuli dichoptically (2IFC of correlated or not, random dots stereograms) and found disparity sensitivity up to ±20 arc min of retinal disparity (Cormack, Scott, & Schor, 1993).

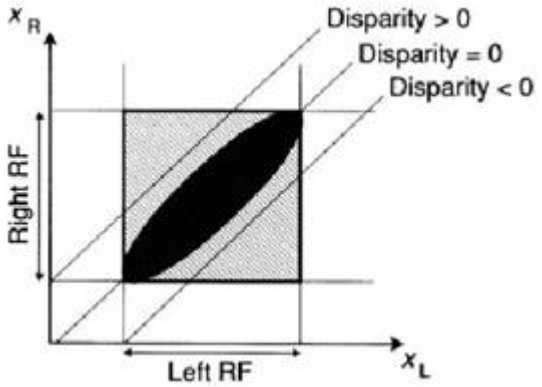


Figure 6 Archetypal binocular disparity detector. A light-bar falling in corresponding retinal points at zero disparity (see Figure 4) will fall on different retinal position in the left or right eye, projecting to correspondent RFs in cortex. Their empirical position is represented on the x and y-axes. In addition, stimuli falling within a specific range of disparities will activate a correspondent pair of cells. An optimal detector would be represented as a square (shaded area in the image), whose main diagonal represents zero disparity. Stimuli falling here will be optimal. Closer or further objects will fall respectively below or above this line, eliciting a proportional reduction in response. Taken from Ohzawa et al. (1990) – copyright license agreement # 4414821457723.

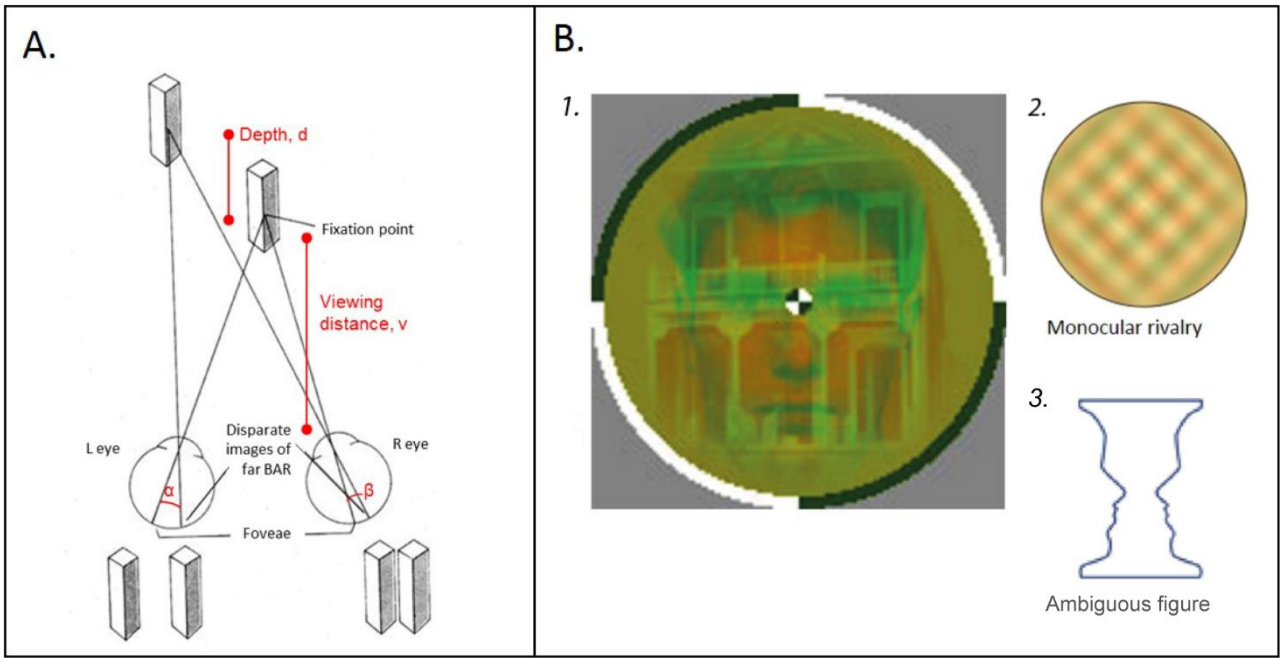


Figure 7 Binocular vision: phenomena **A)** A diagram of binocular crossed-disparity: if a viewer fixates the bar on the right, the far bar (left bar) will generate disparate images on each eye's fovea, respectively at α and β disparities. The amount of disparity is proportional to the viewing distance and the depth between the bars. Resulting retinal images are reproduced flattened, at the bottom. Adapted from Prof. D. Heeger, *"Depth, Size and Shape"*, Department of Psychology, New York University (2006) – copyright permission granted by NYU for 'fair use'. **B)** Three stimuli adopted to induce perceptual alternation (taken from Blake & Logothetis, 2002) – copyright licence agreement # 4414831184177. **1.** Looking through red-green anaglyphs, *binocular rivalry* between a house and a face will occur. **2.** An example of *monocular rivalry*: two superimposed patterns differ in colour and orientation thus resulting in alternating perception; **3.** The Rubin's *ambiguous figure*: appearance fluctuates over time between a vase and two facing human faces.

1.3.1 Stereovision

Stereo vision is the perception of depth based on retinal disparity signal, excluding monocular cues (such as shadows, linear perspective, texture gradient, relative size, etc.).

Look at Figure 7A: when we fixate a point (or the right bar) so that it falls on each eye's fovea, another point (the far left-bar), within the binocular visual field at varying viewing distance (v+d), will fall at different distance from the respective foveae. The resulting discrepancy in retinal-locations (angles α and β), described respectively on the left and the right retina leads to binocular disparity, which can either be "crossed" (α<β) or "uncrossed" (α>β). Cortical cells (in V1 and beyond) are tuned for retinal disparity i.e. they respond maximally to a certain amount of disparity (α –β) (Barlow et al., 1967; Hubel & Wiesel, 1970a).

In addition to retinal disparity, there are also extra-retinal cues to depth. Among others: convergence (the different angles of inclination of the eyes directed to a target), accommodation of the lenses and the refraction power of each eye. The first two are

usually defined as ‘physiological cues to depth’. With normal accommodation, refraction etc. there is a region of visual space within which **binocular fusion** occurs, i.e. the viewer perceives a single image. This region - Panum’s area is thought to be modulated by attention and retinal eccentricity (e.g. fusional space is larger in peripheral retina; Vojnikovic & Tamajo, 2013). Image size also plays a role: the limit for fusion is larger for low SF defined stimuli (broader and blurred) and reduces with increasing SFs (C. Schor, Wood, & Ogawa, 1984). Fusion is also limited by temporal properties of the images: TFs >0.5 c/sec will be less likely to favour a fused perception (C. M. Schor & Tyler, 1981).

When fusion is possible, a problem arises: which are the specific features of an eye’s image that match the other eye’s view (and should therefore be fused)? This question introduces “the correspondence problem”. Stereovision is possible without contours to guide fusion, e.g. when using random dots stereograms (see also 2.3.6, Figure 14A), thus falsifying the view that spatial forms guide fusion. In fact, B. Julesz (1971) stated that stereo images are matched based on (a) physical similarity, (b) one-to-one matching of a features (the uniqueness constraint) and (c) smooth variation of disparity across the target image (the continuity constraint). These features would allow fusion of different monocular image pairs into a correspondent single binocular percept.

As discussed in section 1.3, fusion is possible only within a *gradient* of disparities across space (e.g. for two points: difference in disparities divided by their separation in visual angle). We could think about the gradient as the slant of a surface in space, relative to the point of view of the observer. Fusion will not be achieved if the slant in depth is too steep. Thus, outside Panum’s fusion area the viewer will experience double vision, also called **diplopia**. This phenomenon occurs for example when we try to focus on objects too close to our eyes (i.e. when convergence and accommodation fail).

If different monocular images, eliciting an ambiguous sensory signal, fall on correspondent retinal locations of the two eyes, one’s percept of the stimulus can alternate over time. This phenomenon is known as **rivalry**. Wheatstone, using his stereoscope, investigated the effect of presenting dissimilar alphabetic letters to the two eyes, introducing the phenomenon of <u>binocular rivalry</u> (Wheatstone, 1838). An analogous situation arises under monocular vision, when two stimuli are physically superimposed – e.g. a red and green grating at two different orientation – resulting in our percept alternating between the two components (see Figure 7B.2). Monocular rivalry is related to features that remain visible in the binocular VF but are grouped differently over time, and differs from binocular rivalry where one or the other of the

components disappears (C. Schor et al., 1984). Binocular rivalry can derive from interocular dissimilarities in colour, luminance, contrast polarity, form, size or velocity of the presented images, either embedded in simple stimuli, (e.g. grating of orthogonal orientation) or in complex stimuli (e.g. Figure 7B.1: a house or a face), viewed each in one eye (Blake & Logothetis, 2002). In the presence of a stronger competing-stimulus (e.g. at higher contrast), rivalry will occur quickly and the relative exclusive perceptual dominance will be maintained for longer duration (Blake & Logothetis, 2002).

When two spatially dissimilar images are presented for a sustained period to the two eyes, perceptual dominance could develop: the cortical visual processing will be driven mainly by one eye's stimulus and the stimulus driven by the other eye will be ignored. Thus, monocular **suppression** occurs. This cortical phenomenon is activated by temporal and spatial properties of competing stimuli (P. C. Huang, Baker, & Hess, 2012) and it is commonly refer to as inter-ocular suppression (**IOS**). IOS has been thought to avoid diplopia through "active cortical inhibition of objects in all part of the visual field of one eye" (Jampolsky, 1955). This phenomenon commonly occurs in amblyopia, as will be discussed in section 2.4.2.

A specific case of IOS is defined as **large regional suppression**: children with severe amblyopia (either strabismic or anisometropic amblyopia- see sections 2.1, 2.2) develop a scotoma to eliminate the lack of fusion (due to extreme disparity between retinal images). It is an intermittent phenomenon occurring when vergence movements to solve the severe deviation of one eye are interrupted only occasionally, e.g. due to interfering signal. In this case, the patient will alternate between a state of binocular fusion and one of monocular vision, with concomitant suppression of the input from the deviated eye. We could think of this process as a marker to track the region of space (in the horizontal and vertical dimensions) where the amblyopic viewer is more susceptible to suppress the input to the AE. Details can be found in Wright KW, Spiegel PH, Thompson LS, *"Handbook of Paediatric Strabismus and Amblyopia"* (New York, NY: Springer; 2006), p. 188.

1.4 Development of vision

Much of our understanding of the visual system comes from electrophysiological studies of monkeys. Although broadly similar in structure the human visual system develops at about ¼ of the rate of the monkey i.e. in 1 week after birth a monkey shows similar maturation in visual abilities as a human baby after one month (Teller, 1981). Among

behavioural studies on humans, Frantz, in the early 60s, refined an interest infant's looking-patterns by inventing the method of Preferential Looking (PrL). This allows one to estimate the infants' visual discrimination ability, by measuring his/her spontaneous staring at one of the two patterns. Specific measures used are direction of gaze, duration of fixation and number of distinct stimulus-fixations (Fantz, 1958). A variant on this method is forced-choice preferential looking (FPrL) which involves a 'blinded' adult observer performing discrimination based only on the looking patterns of the baby, allowing one to infer an estimate of the baby's visual ability (Teller, 1979). This method has pushed back the preferred minimum age for vision testing to around 5 months.

In clinical practice, Teller cards are widely used to estimate children's visual function. Cards show gratings at various SFs and by measuring the highest frequency children reliably fixate one can estimate their *acuity* (for norms of use in 1m. to 4y. children see Mayer, Beiser et al., 1995). Other studies used optokinetic nystagmus (OKN) to quantify an oculomotor-reflex (repeated smooth pursuit followed by corrective saccade) in response to a stimulus' motion (Dobson, 1980). Visual Evoked Potentials are a physiological measure used to quantify the cortical response to specific visual stimuli, recording latency and amplitude of neural activity after stimulation (e.g. to a flash light) in a specific state (e.g. wakefulness or sleep)- see Norcia and Tyler (1985). Although highly sensitive, these measurements are time consuming to collect and only rarely reliable in infants under 3 months of age (Barnet, Friedman et al., 1980).

A significant driver for research on infant vision came from the animal studies of cortical plasticity pioneered by Hubel and Wiesel (see section 1.2). Table 2 shows the main milestones the developing visual system achieves within the first 6 months after birth. **Visual acuity** (grating detection) rapidly develops during the first month and progressively improves, reaching 0.7 logMAR around the 6th months after birth. Adults' level of acuity (0 logMAR) will be reached around the age of 3 years (Teller, 1997). A reason for this rapid development was found in kittens, where photoreceptors in the retina centralis and the optic nerve respectively reach 100% and 80% of their maturation by ~8 weeks after birth (the remaining 20% is reached at ~24 months; Ikeda, 1980). Therefore, physiologically, adult-like visual acuity is reached shortly after 24 months of age, but the visual system (particularly parvocellular neurons) is highly susceptible to abnormal development in the preceding period (Ikeda, 1980). **Contrast sensitivity** measures the relative difference in luminance required to support detection of a stimulus. Performance is typically characterised using a Contrast Sensitivity Function (CSF), which plots contrast sensitivity against the spatial frequency of the stimulus

(Figure 8; see also section 2.3.1). Spatial frequency (SF) is a measure of spatial-detail - how often a sinusoidal luminance profile modulates in one degree of visual angle. In infants, the CSF is shifted down compared to normal adults (indicating poorer sensitivity) and extends over a smaller range of SFs (indicating poorer sensitivity to high SFs, i.e. *acuity*). Within few months, the child will achieve sensitivity in detecting gratings at both lower and higher spatial SFs (Banks & Salapatek, 1978).

[the Picture has been removed due to copyright restrictions]

Figure 8 CS curves and development. Average contrast sensitivity (reciprocal of contrast threshold) plotted against increasing spatial frequency (measured in cycles per degree) for children of 1, 2 and 3 months. Data are compared to an adult. Adapted from Banks and Salapatek (1978)

Colour vision develops to an adult level only after the first year, due to a lack of blue-channel activity (short-wavelength-sensitive cones on the retina), likely related to a general lower sensitivity to higher SFs (Teller, 1997). Confronted with 100% contrast stimulus, infants as young as 1 month show an overall lower spatial resolution than adults (factor of 1.5 log units) but no difference in temporal vision (measured by manipulating the temporal frequency (TF) – or flicker rate - of the stimulus). Indeed, their Critical Flicker Frequency (**CFF**) – the highest detectable flicker rate- would be near to adult levels (about 60 Hz) from the age of 2 months (Teller, 1997). **Stereopsis** has been measured using PrL or FPrL of line stereograms or VEP measured using (random dot) coherent motion tasks. No measures of depth perception based on combination of horizontally shifted images has been obtainable before the age of 6 months, implying a late activation of binocular disparity tuned neurons (E. E. Birch, Gwiazda, & Held, 1982). Finally, **vernier acuity**, the ability to detect displacements between two stimuli (e.g. subtle misalignments in near-collinear bars), moves from very poor performance and steeply develops in the first months (Teller, 1997), reaching nearly adult level after 6 months of age (0.05 -1 minutes of arc, i.e. ~1.3-0.0 logMAR).

[the Picture has been removed due to copyright restrictions]

Figure 9 Summary of developmental rates of four resolution abilities: critical flicker frequency (CFF), grating acuity, stereoacuity and vernier acuity. The dashed line corresponds to the estimated performance in adults and marks the 0 point on the y-axis. The abilities are plotted from birth to 12 months of age (x-axis) in logarithmic units of proportional decrement from adults' performance (y-axis). From Teller (1997)

The tasks discussed so far -orientation selectivity, SF tuning etc. – are thought to be mediated by low-level visual mechanisms e.g. in primary visual cortex. Recently there has been more consideration of the development of extra-striate processing, responsible for segmentation, integrative motion and pattern processing, in order to refine our knowledge on normal development and the prevention/treatment of abnormal development (for a review see Braddick & Atkinson, 2011).

Table 2 summarises the commonly accepted knowledge of the development of various visual abilities from birth. Note that there is evidence of individual differences in rate of visual development (Braddick & Atkinson, 2011), thus Table 2 provides only an indicative summary.

	Newborn	@ 1 m.	@ 2 m.	@ 3 m.	@ 6 m.	Adult norm
Acuity * (logMAR)	≃1 c/deg (1.5)	1.5-2.4c/deg (1.3-1.1)	2.8c/deg (1.03)	3-4c/deg (1-0.9)	6.4c/deg (0.7)	30c/deg (0)
Contrast Sensitivity peak#	/	9 (11.1%)	11-19 (9.09-5.26%)	/	/	500 (0.2%)
Motion discrim.	Very poor (only lowSF-high C)	Gross	Improving	/	excellent	excellent
Critical Flicker Frequency	Not obt.	≃40Hz	≃50Hz	≃52Hz	55Hz	55-60 Hz
Stereo –acuity (arc units)	absent	absent	absent	≃60 min	1 min	30-55 sec
Colour vision	absent	absent	rudimentary (L,Mcones)	better	Better (+Scones)	L,M,S cones

* Acuity is defined as the highest SF grating visible at maximum contrast, and it is expressed in cycles per degree (c/deg) or as the corresponding logarithm of the minimum angle of resolution (logMAR)
# The temporal contrast sensitivity function is shifted by 1.5 log units in children of 1 month of age compared to adults, but has a nearly identical shape.

Table 2 Summary of visual development in babies, comparing performance at 1, 2, 3, and 6 months to adult norm. Results were obtained using behavioural and VEP measures, typically with grating stimuli (spatial and temporal resolution and colour vision) or stereo-lines (stereoacuity). From: Teller (1997)-motion CFF, stereo and colour vision information, Braddick and Atkinson (2011)-for a general comparison and contrast values, and Banks and Salapatek (1978)-for acuity details.

With this snapshot of normal visual development in mind, we now move onto consider abnormal visual development and in particular amblyopia, the focus of this project.

2 Amblyopia: introduction

Amblyopia is defined as unilateral (rarely bilateral) reduction of the best-corrected visual acuity (BCVA) in an otherwise healthy eye (Ciuffreda, Levi, & Selenow, 1991), i.e. in the affected/amblyopic eye (AE). This condition originates from prolonged abnormal retinal stimulation during the development of the visual system. According to its origin (also called, the amblyogenic factor), amblyopia is defined as *strabismic* (due to persistent inter-ocular misalignment), *refractive* (by a refractive imbalance) or *visual deprivation* amblyopia (due to congenital monocular obstruction, i.e. ptosis or cataract) (American Academy of Ophthalmology & Pediatric Opthalmology/Strabismus Panel, 2012). Specifically, refractive amblyopia is re-defined as *anisometropic* - when the deficit is unilateral, or, less commonly, as *ametropic* or *meridional* - after a bilateral deficit, i.e. respectively either a significant refractive error or astigmatism, in both eyes. When strabismus and anisometropia co-occur, amblyopia is defined *combined-mechanism* (Cotter et al., 2012), or also *mixed* (Stewart, Fielder et al., 2005). Finally, there is a specific form of deprivation amblyopia that may occur as adverse effect of patching therapy: it has been recently defined as *occlusion* or *reverse* amblyopia (David K. Wallace, Repka et al., 2018). In this thesis, we will focus on anisometropic, strabismic and combined-mechanism amblyopia (unilateral amblyopia only).

Clinically, in the presence of at least one amblyogenic factor, unilateral amblyopia is diagnosed based on the corrected visual acuity. The specific criteria have changed towards the end of the last century: from a BCVA in the AE equal or worse than [0.2-0.3] logMAR being sufficient to diagnose amblyopia, also a persistent inter-ocular difference in best-corrected acuity (IOAD; usually, ≥0.2 logMAR) was then required for the same diagnosis (J. M. Holmes & Clarke, 2006). Some research-groups define amblyopia based on IOAD >0.1 logMAR (MOTAS, ROTAS groups; e.g. Stewart, Stephens et al., 2007), which is enough to exceed the acuity test-retest variability (J. M. Holmes, Beck et al., 2001), while others accept only IOAD >0.2 logMAR (PEDIG group; e.g. J. M. Holmes, Kraker et al., 2003; Repka, Beck et al., 2003). Proviso a minimum level of IOAD, depending on the BCVA in the AE, amblyopia is classified as *moderate* (BCVA in the AE: ≥0.3 to ≤0.6 logMAR) or *severe* (BCVA in AE: ≥0.7 logMAR). Note that acuity measures >1.3logMAR are likely related to ocular conditions other than

unilateral amblyopia, therefore additional checks are highly recommended (David K. Wallace et al., 2018). *Mild* amblyopia corresponds to a BCVA in the AE of 0.1 to 0.3 logMAR (Stewart et al., 2005) - thus =0.3logMAR is associated with either mild or moderate amblyopia, depending on the study group.
In this thesis, we will associate amblyopia to IOAD≥0.2 logMAR and define the deficit as mild (BCVA in the AE ≤0.2logMAR), moderate (BCVA in the AE ≥0.3 & ≤0.6) or severe (BCVA in the AE ≥0.7logMAR).

Amblyopia is one of the leading causes of visual loss in children with an estimated prevalence of ~3% (Attebo, Mitchell et al., 1998; J. R. Thompson, Woodruff et al., 1991), varying from 0.7 to 5.5 % in population-based studies, depending on the population investigated and the definition used for amblyopia (David K. Wallace et al., 2018; see Table 3 for comparable studies in children aged 6-72 months). Bilateral amblyopia (either due to ametropia or bilateral vision deprivation) is much less frequent than unilateral (e.g. see Table 3), although the exact proportion varies between studies. Henceforth, we will refer to *unilateral* amblyopia, unless differently specified.

In a month, amblyopia can account for >75% of outpatients visits to NHS paediatric eye services (Stewart, Shah et al., 2016). It is an important public health problem, associated with around 2.6 relative risk for lifetime binocular visual impairment (Chua & Mitchell, 2004; van Leeuwen, Eijkemans et al., 2007). Amblyopia and/or its treatment can cause distress in young patients and/or parents (Hrisos, Clarke, & Wright, 2004) and have a negative impact on their quality of life, for example by affecting family life or social interactions (Carlton & Kaltenthaler, 2011). Untreated or residual amblyopia can also have a negative impact on adults' education and/or occupation (Chua & Mitchell, 2004).

Monocular amblyogenic factors lead to degradation in image quality, which in turn leads to (a) reduced reliance by the patient on the AE, aka interocular suppression (IOS), and (b) the syndrome of functional visual deficits termed *amblyopia*. As said, clinically, amblyopia is primarily associated with reduced visual acuity (Ciuffreda et al., 1991), but the syndrome also includes poor contrast sensitivity (Levi & Harwerth, 1977), elevated foveal "crowding" (the interfering effect of visual "clutter" that ordinarily is only manifest in the peripheral visual field; Levi & Klein, 1983), fixation instability (K. R. Kelly, Jost et al., 2015), and poor or absent stereopsis (Weakley, 2001). In particular, anisometropic amblyopia, compared to strabismic, is associated with a more moderate loss in visual acuity, poorer contrast sensitivity especially at high spatial frequencies and retention of a certain level of binocular visual function (McKee, Levi, & Movshon, 2003).

2.1 Anisometropia

Around 20% of the adult population exhibits a degree of *anisometropia*, - an IOD in spherical-equivalent refraction (SE; i.e. sphere + cylinder/2) of a certain magnitude, e.g. ~15% adults: IOD≥1D (D, dioptres= 1/focal length), commonly due to a difference in axial length and/or optical power (i.e. corneal and lens power in redirecting light-input) between the eyes - for a review see Barrett, Bradley, and Candy (2013). The consequence may be myopia (~ 60-70%; near-sighted: focal point before the retina) or hyperopia (farsighted: focal point beyond the retina). An additional correction (for astigmatism) will be prescribed if the refractive difference occurs at any meridian (cylinder axis). The interocular difference in refractive power (anisometropia) can be accompanied by ***aniseikonia***, a difference in magnification between the two eyes causing the images to differ in their perceived size (i.e. AE image: stretched in case of myopia, compressed in case of hyperopia) with consequent abnormal sensitivity over a full range of SFs (Bradley & Freeman, 1981). Since long ago, a degree of aniseikonia greater or equal to 0.75% size difference between monocular retinal images has been considered clinically significant when associated with constant eye strain and/or headache, not relieved by accurate refractive or motility corrections (Burian, 1943). As a rule of thumb, the expected relative difference in image size is about 1% per dioptre of anisometropia, with the limit of tolerance in humans commonly set at >5% image size difference (Achiron, Witkin et al., 1997). Exceeding this limit, prevents acquisition of fine binocular vision, although in some cases a higher degree of aniseikonia can still be tolerated by anisometropes, who maintain eye alignment: some might pass Random Dot Stereogram tests although their binocular vision is not optimal (Campos & Enoch, 1980). Caution should be taken when using just one stereo-test for screening binocularity, especially in presence of aniseikonia.

Around 50-75% of cases of amblyopia are associated with some degree of anisometropia (together with strabismus in ~10-30% cases) and hyperopia is associated with amblyopia more frequently than myopia, with a risk estimated to be twice as high (Levi, McKee, & Movshon, 2011). Such statistics may be influenced by the reliance of screening procedures on near-vision tasks which are more sensitive to hyperopia (Attebo et al., 1998). For details on different populations – see Table 3.

In about one third of cases, anisometropic amblyopia is likely to resolve after a period of optical correction alone, usually ranging 2 to 4 months to reach a plateau, depending on the baseline vision (Cotter, Edwards et al., 2006; Moseley, Fielder, & Stewart, 2009). Alternatively, a refractive-surgery intervention might be recommended in children who cannot tolerate spectacles or contact lenses (W. F. Astle, Fawcett et al., 2008). However, surgery is widely approved only in adults above the age of 18 yrs, as refractive errors are less stable in children and, especially in young children, response to surgery is less predictable and use of general anaesthesia carries risk. Refractive surgery include intraocular techniques, such as implantation of an intraocular lens (if missing, i.e. in case of aphakia) or removal of the crystalline lens (in case of high unilateral myopia) & extraocular techniques, i.e. corneal surgery methods (in case of unilateral myopia, astigmatism, and hypermetropia), including photorefractive keratectomy and laser-assisted procedures (Alio, Wolter et al., 2011).

2.2 Strabismus

Commonly called “squint”, this ocular misalignment is routinely diagnosed using either a cover-uncover test, to evaluate visual-axes direction and/or a prism cover test (PCT- using an accommodative target, at 6m for distance and 40cm for near), to measure the angle of deviation (in Prism Dioptres, PD or Δ; with and without wearing glasses). Ocular movements (version, duction) including motor fusion, and head-posture, are normally tested in all positions of gaze to check for anomalies. Strabismus is defined as manifest (-tropia) or latent (-phoria) monocular deviation (horizontal: eso-, convergent to nasal septum OR exo-, divergent from it; vertical: hyper- OR hypo-) when respectively occurring during binocular vision or after its interruption (covering the non-deviating eye). In both cases, a certain magnitude (Δ) of deviation (mild if <10Δ) can be constant or intermittent, i.e. with or without constant monocular fixation. “Accommodative” indicates that the deviation is reduced or corrected by wearing the appropriate prescription. It is then broken down into *constant accom.-* or *-with accom. element* OR *fully accom.-.* In these cases, abnormal retinal correspondence (ARC) can develop if a small angle of deviation is present – see section 1.3. Also, strabismus can be defined *incomitant* when the deviation varies depending on the position of gaze, or *concomitant*, when independent to gaze position (more typical of developmental anomalies). To find more see “*Guidelines for the Management of Strabismus in Childhood”* (The Royal College of Ophthalmologists; 2012).

The prevalence of strabismus in children aged 6 to 72 months varies from 2.1% to 3.3% depending on region, with most studies showing results around 3%, and further amblyopia occurs in ≈1.7% on average (see Table 3).

Study	Ethnicity	Prevalence (%)			
		Strab.	Aniso.	Amblyopia (bilateral)	
MPEDS Group * (MEPEDS, 2008) **(Borchert, Tarczy-Hornoch et al., 2010)	Hispanic	2.4*	4.3**	2.6 (0.5*)	74% pure aniso
	Afro-American	2.5*	4.2**	1.5 (0.4*)	
BPEDS Group § (Friedman, Repka et al., 2009) §§ (Giordano, Friedman et al., 2009)	non-Hispanic	3.3§ (50%ET)	5.0§§	1.8 (0.0§)	32% pure aniso
	Afro-American	2.1§ (50%ET)	4.3§§	0.8 (0.1§)	

Table 3 Population studies in children 6 to 72 months to assess the prevalence of anisometropia, strabismus and amblyopia. Respectively: anisometropia (aniso.)=at least in one eye: ≥1D SE hyperopia or ≥3D SE myopia or ≥1.5D astigmatism; strabismus (strab.)=at least in one eye: constant or intermittent tropia). Amblyopia defined as ≥0.2logMAR IOAD and VA in the AE≥0.2 logMAR) –in parenthesis, the specific proportion of bilateral amblyopia (i.e. in both eyes: BCVA>0.3logMAR with either obstruction of the visual axis or ametropia - hyperopia ≥4D SE, myopia ≥6D SE, or astigmatism≥1.5D). Specific studies are indicated in column 1, left; MPEDS=Multi-Ethnic Paediatric Eye Disease Study; BPEDS=Baltimore Paediatric Eye Disease Study Each respective investigated group is specified in column2 (ethnicity), and the relative findings are reported in columns 3 to 5 (prevalence; in order of: strabismus, anisometropia and amblyopia). ET=esotropia; SE=spherical equivalent; D=dioptre.

2.3 The functional deficit in amblyopia

Psychophysical studies have played a fundamental role in defining the functional deficit associated with human amblyopia and the initial behavioural analysis concentrated on monocular abilities – particularly contrast sensitivity and visual acuity.

2.3.1 Visual contrast

Visual contrast (introduced in section 1.4) is the measure of the range of luminance (*L*) present in a visual stimulus (independent from absolute luminance, i.e. light intensity on a surface, in cd/m^2; size varying with grating's SF). Different definitions of contrast exist whose appropriateness depends to a degree on the nature of the stimulus. The contrast of periodic stimuli (e.g. sine-wave gratings – see Figure 10) is often expressed using Michelson contrast: $C = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}$. For aperiodic stimuli, e.g. uniform patches against a

background, contrast is often quantified using Weber contrast: $C = \frac{L_{\max} - L_{\min}}{L_{\text{background}}}$. The contrast of more complex aperiodic stimuli such as natural images is usually expressed in units of 'Root Mean Square' contrast: $C = \frac{L_\sigma}{L_\mu}$, where L_σ=L standard deviation, L_μ=mean L.

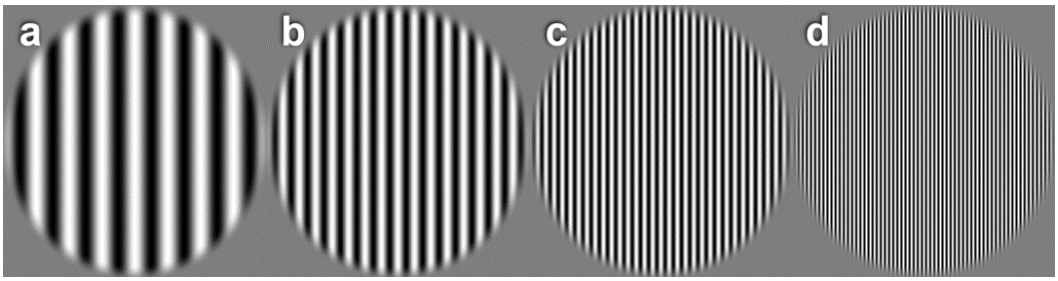


Figure 10 Example of grating stimuli from (left) low to (right) high spatial frequency (SF). Stimuli are shown at maximum contrast spanning three octaves of SF, here (**a**) 8, (**b**) 16, (**c**) 32 and (**d**) 64 cycles of the luminance-defined sine-wave per image. SF is normally expressed in cycles per degree of visual angle (c/deg).

Contrast-sensitivity (CS) is defined as the inverse of contrast-detection-threshold, the minimum contrast necessary to elicit a criterion level (e.g. 75% correct) of detection. High contrast thresholds indicate poor performance while high CS indicates good performance. CS-plot as a function of stimulus spatial frequency (SF; the rate of luminance-modulation in space), is known as the Contrast Sensitivity Function (CSF). This curve defines the limits of visibility of a sine-wave stimulus as a function of its spatial frequency and typically is described as having an inverted U-shape (Campbell & Robson, 1968). Adults' performance is best at 4-5 c/deg at near viewing distance, with the limits of visibility falling between around 0.1 to 30 c/deg. *Visual Acuity (VA)* is a measure of precision in spatial resolution, to identify the smallest visible letters or optotypes, presented at high contrast –suprathreshold- (measured using different units, e.g. logMAR: logarithm of the Minimum Angle of Resolution, or the visual angle subtended by the target). Similarly, Grating VA is a measure of the limit of visibility of sine-wave gratings, as the ones in Figure 10, defined by contrast (amplitude) and SF (bandwidth).

Observers with anisometropic and strabismic amblyopia show poorer contrast sensitivity, particularly at higher SFs. Figure 11 (right) shows that the CSF for an anisometropic observers' AE, peaks at a lower SF and exhibits lower overall sensitivity than when measured using the fellow eye (FE). For such observers, it is generally assumed that poor focus of the affected eye leads to under-representation (or

dysfunction) of cells tuned to high spatial frequencies (Bradley & Freeman, 1981; Levi et al., 2011). Plus, uncorrected aniseikonia could account for significant loss of sensitivity at low SFs, as the perceived size difference (abnormal magnification of the image seen by the AE) tended to enhance sensitivity to details, i.e. to high SFs, so selectively affecting sensitivity, i.e. greater IOD to low SFs (Bradley & Freeman, 1981). Figure 11 (left) shows that the loss in strabismic amblyopia is less marked and limited to detection of high SF stimuli. Sensitivity in strabismus is markedly influenced by stimulus configuration and thus elevated “crowding” may contribute to poor sensitivity at high SFs which contain more repeating/periodic structure (Levi & Klein, 1983). However, the perceptual loss in these amblyopes is less clear and sometimes performance is poor at any SFs (R. F. Hess & Howell, 1977) so that a specific evaluation of contrast sensitivity should be necessary to prescribe an appropriate treatment (e.g. a defocused image would not work in presence of a low frequency abnormality too).

[the Picture has been removed due to copyright restrictions]

Figure 11 Examples of contrast sensitivity functions from amblyopic patients with strabismus (on the left; adapted from R. F. Hess & Howell, 1977) or anisometropia (on the right; adapted from Levi & Harwerth, 1977). On each graph, sensitivity in FE (the upper curve) is normal. AE results less sensitive in both amblyopes, with strabismus responsible of significant loss only at high SFs (left; greater loss above ≈4 c/deg); while anisometropia significantly reduces overall AE’s sensitivity (right). In the top left corner, an illustration of the contrast sensitivity function in adults (≈2-30 c/deg; log steps; against mid grey background), firstly presented by Campbell, Robson 1968.

McKee and colleagues tested a large number of adults with amblyopia to examine the overall deficit on spatial vision, to overcome case-studies differences, and to refine **specific patterns** for type *and* severity of the deficit in amblyopia (McKee et al., 2003). They tested 427 adults with amblyopia (plus 68 controls) measuring monocular spatial resolution: (i) visual acuity, using Snellen optotypes, vernier alignment, contrast-gratings, and (ii) contrast sensitivity, using the Pelli-Robson chart (1988; contrast-defined-letter chart, identified at 1m. viewing distance) and edge sensitivity test (luminance step detection, scaled for grating acuity). Furthermore, they investigated the

(iii) presence of binocularity – using a motion integration test (discrimination of direction of a phase-shift: 2Hz horizontal gratings dichoptically presented 90deg out-of-phase, at increasing SFs) and stereo-optical-circles (coherent motion discrimination of global-shape). Anisometropic amblyopia was associated with poor acuity (no significant differences at different tests), poor contrast sensitivity (especially at high SFs) and a residual level of binocularity. Strabismic amblyopia, instead, was associated with poorer acuity than anisometropic amblyopia (particularly optotype and vernier acuities) and relatively spared grating-contrast sensitivity. However, binocularity was severely impaired. The authors concluded that VA loss (optotype acuity) may predict the amount of amblyopia-deficit but the presence of binocularity is also critical. Indeed, analysis based on binocularity showed that a significant degree of binocularity-*loss* accounted for closer to normal CSF but poorer VA (optotypes, vernier) at any SFs (measured at grating-VA) compared to amblyopes with a *residual*-binocularity. In summary, mild-to-moderate anisometropic amblyopia (AE optotype acuity from 0.3 to 0.6 logMAR) is characterized by a greater loss of contrast sensitivity and maintenance of binocularity, while in mild-to-moderate strabismic amblyopia CS is less affected but the condition significantly affects binocularity. Thus, more sensitive measures of binocularity would be useful in the differentiating subtypes of amblyopia. To this end, Kwon et al (2015) propose a measure of binocularity based on *SF dependent-contrast imbalance,* which uses dichoptic presentation of SF-band-pass letter-pairs to establish the intra-ocular contrast-difference required for letters to be equally likely to be perceptually dominant.

We now consider how visual abilities, other than contrast sensitivity and acuity, are affected in amblyopia.

2.3.2 Positional coding

Positional acuity (precision in judging the relative location of features) has been investigated in amblyopic viewers using vernier-alignment resolution of flanked targets (Levi & Klein, 1985), bisection judgements (Levi & Klein, 1983), or judgements of positional jitter (Levi, Klein, & Sharma, 1999). Anisometropic amblyopia does not affect performance on either vernier or bisection tasks (foveal thresholds for spacing is close to normal - 5-10arc sec - when scaling the stimuli presented to AE to compensate for resolution loss), while strabismus causes a severe loss of spatial resolution in the same conditions, both in detection and discrimination tasks (McKee et al., 2003). The

tolerance of positional-jitter (in orientation detection task of suprathreshold Gabor[3] patches) is affected in human amblyopia: strabismus causes a loss in sensitivity for foveal presentations comparable with similar deficit measured in periphery in subjects with anisometropic amblyopia (Levi & Klein, 1985). This effect may arise from abnormal cortical activity as (i) a consequence of *under-sampling* (fewer neurons and RFs driven by AE) of the retinal image from the AE or (ii) as a spatial disarray in RFs location (distorted topographic mapping in correspondent RFs) leading to a raised *positional uncertainty*. Neither of these mechanisms alone can explain amblyopia without considering the hypothesis of higher neural noise in V1 (Levi et al., 1999).

2.3.3 Spatial distortion

Since the late 1950s, there has been evidence that amblyopia induces abnormalities in perceived spacing, segmentation and warping of Snellen letters when viewing stimuli through their AE (Pugh, 1958). This is influenced by the pattern configuration (e.g. proximity, shape) and was initially thought to be caused by fixation instability. The presence of strabismus was found to affect monocular spatial vision, causing errors in vertical alignment and horizontal bisection (Bedell & Flom, 1981). Strabismic amblyopes reported perceived compression and expansion of horizontal spatial relations when either judging the alignment of circle or attempting to judge their midpoint using their AE (Sireteanu, Thiel et al., 2008). In the same study, observers with anisometropic amblyopia were more uncertain (i.e. higher positional jitter in subjective point-to-point reconstruction), but exhibited less systematic distortion (i.e. limited biases in perceived position). In general, greater errors (uncertainty and distortion) were associated with severity of amblyopia and errors were more evident when stimuli were presented across sensory modalities (radial displacement plus acoustic cues to indicate the angular position). Further, errors were particularly marked for observers with no residual binocularity and large constant deviation. The notion is then that resultant poor binocularity compromises positional mapping of RFs (Sireteanu et al., 2008). In a dichoptic localization task - where a cross-hair cursor is only visible to the AE and has to be placed over a target presented to the FE - the majority of adult amblyopes with strabismus showed greater distortion than in the non-dichoptic condition, whereas anisometropic amblyopes were less affected by modality presentation (M. E. F. Piano, Bex, & Simmers, 2015). These results suggest that higher distortions and uncertainty in

[3] A *Gabor* patch is a sine-wave grating (carrier) windowed by a two-dimensional Gaussian envelope. Gabors are widely used in vision research as they match the receptive field properties of cells in the visual cortex and allow for selective presentation of information at different SFs and orientations.

spatial localisation are mainly influenced by poor binocularity (poor fusion and stereopsis) and a large angle of deviation. They are inconsistent with the idea that poor acuity directly leads to spatial distortions (Bedell & Flom, 1981; Fronius, Sireteanu, & Zubcov, 2004; Sireteanu et al., 2008).

[the Picture has been removed due to copyright restrictions]

Figure 12 Glass patterns (Glass, 1969) are commonly used to identify a deficit in global from perception. The figure shows an example of a concentric pattern: the global form is only perceived when local paired random dots (here at 80% coherence) are integrated. Taken from Rislove, Hall et al. (2010)

2.3.4 Global form perception

Amblyopic viewers show poor perception of global form: spatial structure defined by a series of local features (e.g. lines, dots, etc.; an example in Figure 12). This type of task is interesting because the current view of form processing is that it is hierarchical, with local features (e.g. SF, orientation, motion) being first signalled by neurons with small receptive fields (e.g. within V1) and these signals then being pooled/bound across space by the operation of neurons with larger receptive fields (e.g. within V2/V4). Accordingly, reduced spatial resolution, contrast sensitivity and lower number of activated binocular cells could explain the deficit in amblyopia (thus related to local estimates of image-structure). However, as explained below, also the ability to group local features over an extended area of visual space (integrative visual processing, beyond V1) would be impaired in amblyopia. This would indicate a deficit in global perception.

We already encountered global form when describing the ability to discriminate jittered E-like patterns: strabismic-observers needed more elements to identify the target – visible only if individual elements (jittered Gabors) were combined across space (Levi et al., 1999).

Contour integration also requires to process global form, and specifically strabismic amblyopes showed reduced sensitivity when asked to discriminate which frame

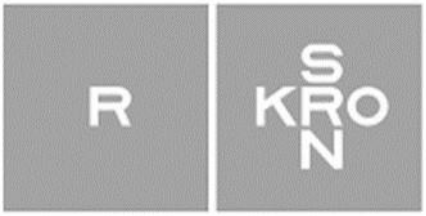
contained a path of Gabor patches (with varying levels of jitter added to the path element angle) against randomly oriented Gabors (R. F. Hess, McIlhagga, & Field, 1997). The authors argued that only “intrinsic positional uncertainty” could explain how each eye’s view was differently affected: low variance in positional jitter was less disruptive to performance when measured through AE compared to the FE, showing higher uncertainty of the AE in positional encoding, i.e. RFs disarray (R. F. Hess et al., 1997). Indeed, anisometropic amblyopes (whose positional uncertainty is usually normal) showed no significant difference in detecting Gabor-paths in noise when using either eye (R. F. Hess & Demanins, 1998). Using an averaging paradigm (what is the signalled overall direction of orientation?) all amblyopic observers were able to discriminate the global average-orientation of systematically varied arrays of Gabor patches, while using either FE or AE, i.e. simple integration was only partially affected compared to normal. However, under dichoptic presentation (in the context of a perceptual matching task) the perceived variance of global orientation and position was imbalanced between eyes (higher level of variance needed in FE to match the variability perceived in AE), suggesting a generalised higher spatial uncertainty in global form processing in amblyopic vs. normal viewers (A. J. Simmers & Bex, 2004)

The amblyopic deficit in *global* perception has been examined using a variety of tasks, comparing poor performance at making local estimates (i.e. higher local uncertainty) to the inability to integrate information across space (i.e. poor global processing). Abnormalities have generally been found in amblyopia for global-orientation processing, e.g. average orientation of locally jittered signal within a Gabors array (Husk & Hess, 2013), global motion processing e.g. coherence[4] task (A. J. Simmers, Ledgeway et al., 2003) and structure-from-motion e.g. discrimination of same/different shapes defined by perturbing their elements: motion defined depth-cues (Husk, Farivar, & Hess, 2012). In addition, detection and discrimination of biological motion perception is known to be impaired in amblyopia, e.g. masking AE was less effective on performance at all levels of coherence in point-light “walkers” (B. Thompson, Troje et al., 2008) . In addition, visuo-motor coordination is abnormal in amblyopia. For example, latency (reaction time and acceleration of reaching movement) and (to a lesser extent) precision (the variable error between trials) of reach-to-touch movements towards lateral stimuli under AE

[4] In motion coherence paradigms observers are required to discriminate the direction of a pattern composed of moving random dots. The pattern is composed of two populations of dots: signal-dots are moving in the same (coherent) direction, noise-dots are moving in random directions. By manipulating the proportion of signal to noise dots one can determine the minimum number of signal dots required to support reliable direction discrimination. This is known as the *motion coherence threshold*. Such tasks have also been adapted to study orientation, size and other visual attributes.

viewing are both affected in both anisometropic (Niechwiej-Szwedo, Goltz et al., 2011) and strabismic amblyopia (Niechwiej-Szwedo, Goltz et al., 2014). Accuracy (constant error) in both types of amblyopia is comparable to normal controls.

Recently it has been reported that face perception is compromised in strabismic amblyopic viewers, with observers exhibiting poor detection and discrimination of changes in facial configuration (relation between features) but not of single features (Cattaneo, Vecchi et al., 2013).



Unflanked *Flanked*

Figure 13 Example of stimuli used to measure acuity at varying sizes and spacing. Single letter presentation (left) and/or crowded presentation (right) are used to obtain unflanked and flanked acuity respectively. For a study using similar stimuli see Song et al., 2014.

2.3.5 Crowding

Crowding refers to the disruptive influence of clutter on object recognition (Levi, 2008). For normal observers crowding only impacts recognition of objects (e.g. letters, in Figure 13) when they are presented in the periphery. In contrast the disruptive effect of context on foveal object recognition is thought to be a consequence of masking, or the perceptual suppression induced by spatial interactions target-flankers (Levi, Klein, & Hariharan, 2002). In amblyopia crowding also affects central vision (Levi & Klein, 1983). In order for acuity testing in children to be effective at detecting the presence of amblyopia, it is recommended that they be tested with crowded optotypes (Simons, 1983).

Levi and Klein (1985) showed that spatial offset discrimination thresholds varied proportionally to target-to-flanks spacing in both controls and amblyopic observers when using stimuli scaled for un-flanked acuity (AE view resulting in higher thresholds). Only strabismic observers also showed abnormalities when using their FE. When stimuli were presented in the periphery, strabismic amblyopes' vernier acuity was proportionally worse than grating acuity compared to controls and anisometropes. For foveal presentation, strabismics were poor at both tasks compared to controls and anisometropes, and indeed their performance was comparable to control subjects'

performance under peripheral viewing (Levi & Klein, 1985). Song *et al* (2014) showed that crowding in anisometropia and strabismus could produce different patterns of performance. Specifically, they showed that the ratio of threshold spacing (between target and flanker) could differentiate the two conditions (see Figure 13 for an example of their stimuli). Further, strabismic amblyopia results in *temporal-crowding* - effect of clutter nearby in *time*, e.g. within rapid serial visual presentation; as well as in *spatial-crowding* - effect of clutter in space, e.g. crowded tumbling-E chart in the presence of similar flankers (Bonneh, Sagi, & Polat, 2007). These different patterns of crowding between normal and amblyopic viewers and between types of amblyopia may relate to the amount of binocularity loss (Song et al., 2014), as a limited correspondence between cyclopean images (ultimately preventing stereopsis) would be more disruptive in the presence of 'more details to match' (as in a crowded visual scene). Indeed, the inter-ocular difference in crowded acuity (measured with minimum distance target-flankers) is associated with the severity of acuity loss (IOAD), in amblyopic and non-amblyopic children (Greenwood, Tailor et al., 2012). This foveal deficit in acuity (crowded-distance and size) is also associated with poor or absent binocularity (stereopsis measured with random-dots stereograms) in strabismic and mixed amblyopia (Greenwood et al., 2012).

2.3.6 Stereopsis *(depth perception)*

A common way to evaluate binocular vision in the clinic is to measure stereopsis (a term introduced by Wheatstone, 1838). As explained in section 1.3, this is the ability to evaluate depth and estimate relative distance between objects in the binocular visual field based on the horizontal disparity between correspondent retinal images (Ohzawa et al., 1990). Since our eyes are offset horizontally, the two retinal images will be shifted in relation to one other, with a single point in space stimulating *corresponding* retinal loci (see Figure 14B). This provides a disparity signal to cortical cells that supports stereo-perception – see section 1.3-Depth/Disparity tuning. To assess stereopsis multiple tests have been developed which generally require the observer to differentiate a form/figure from background after fusing stereo-images, e.g. Figure 14A. An object falling closer to the viewer than the horopter will have disparity<0 (see Figure 14B) and we need to cross our eyes to fixate on it; conversely for disparities >0, uncrossed movement will be necessary. Vergence movement of the eyes (crossed/ uncrossed) and accommodation of the lens allow for retinal correspondence of each eye's image (and also provide independent extra-retinal cues to depth). One could try to fuse the boxes in Figure 14A top (i.e. combine them by crossing the eyes, at ~arm-length viewing distance). Each box (right and left squares) contains a same portion of dots shifted horizontally, in opposite direction to the other (slots left over filled with random dots). Using the depth cue elicited

by the disparity of retinal images corresponding to these sub-regions, the observer who has fused successfully should perceive a square floating in front of a background. The amount of retinal disparity (in this case the horizontal offset of the sub-regions) required for an observer to perceive depth determines his/her level of stereopsis (the finer the stereopsis, the smaller the horizontal displacement of stereograms).

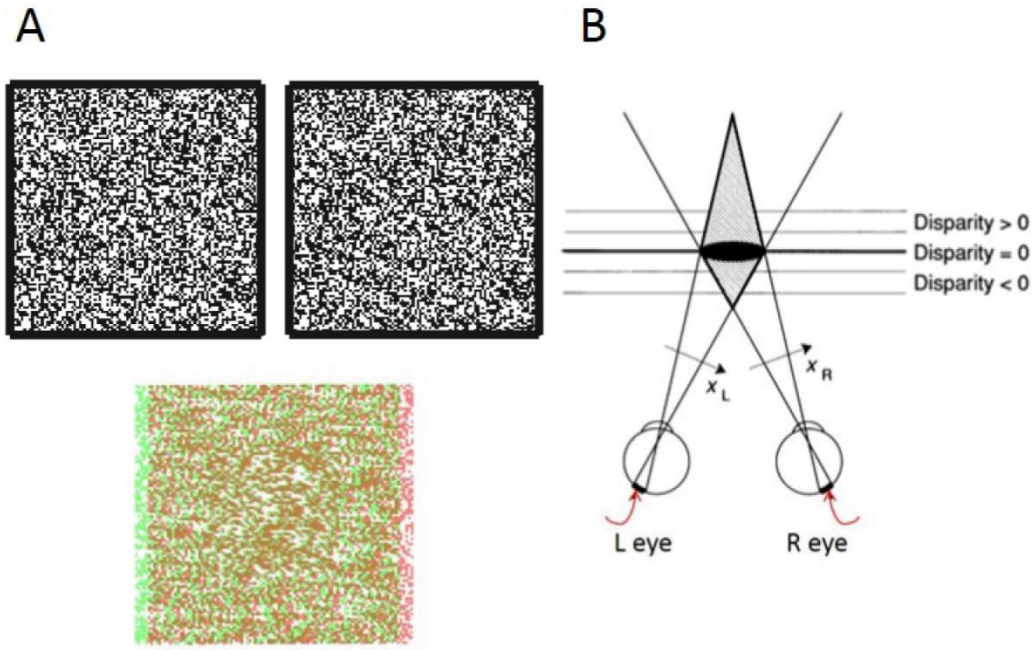


Figure 14 A) Random-Dot stereograms used to measure depth perception. Images are presented independently to the left and right eyes. Each image is composed of a series of dots at identical locations, apart from a sub region of each where the dots are laterally displaced (across the two eyes). This arrangement leads to impression of depth when the two images are perceptually fused. **Top:** random-dot patterns designed to be *free fused* (adapted from Fricke & Siderov, 1997). The observer relaxes accommodation to fuse the two images. Here, a square should appear. **Below:** a random-dot stereogram designed to be viewed through red-green anaglyph glasses (originally introduced by Julesz, 1971): the observer should perceive a floating central square **B)** Geometry of binocular viewing: horizontal position on left and right retina (x_L or x_R respectively) are computed as distance from the focal point, here falling in the middle of the retinal-projections, as indicated by the red arrows. Adapted from Ohzawa et al. (1990) -– copyright license agreement # 4414821457723.

When strabismus is present, it will rarely be possible to obtain a measure of stereoacuity. Currently it is assumed this arises from uncorrelated binocular stimulation preventing development of stereopsis in the first 6 months of life (Braddick, 1996). See also sections 1.3 and 1.4. Further most subjects with strabismic amblyopia fail stereoacuity tests, e.g. in McKee et al. (2003), only 10% of pure or mixed strabismics

passed tests of binocularity (specifically, a binocular motion integration-BMI task[5] and the stereo-circles test[6]).

Overall, and perhaps counter intuitively, while poor binocularity (related to strabismus) is associated with poor acuity it has been linked to higher contrast sensitivity, while the presence of a degree of binocularity (associated with anisometropia) is associated with better acuity and poorer contrast sensitivity (McKee et al., 2003).

With respect to the mechanism by which this pattern might emerge, for anisometropia, defocused but correlated binocular stimulation would allow stereoacuity to develop to a certain degree whereas the binocular de-correlation resulting from strabismus is less likely to lead to development of binocularity. As to how binocularity might impair CS, McKee et al. (2003) speculate that an absence of binocular function could be associated with an over-representation of monocular cells which would confer superior monocular contrast sensitivity. They also propose that superior acuity associated with better binocularity arises from a superior ability to allocate attentional resource to the tested eye. It turned out that abnormal hyperacuity (monocular) in amblyopes can be predicted by the residual level of binocularity. A strong correlation was indeed found between binocularity (tested at stereo tests, including BMI) and vernier acuity results (Agrawal, Conner et al., 2006).

Another way of testing binocular vision is to modulate each eye's input to find out how much each one weights in the binocular percept, i.e. in which condition one eye's input 'stops' the other from being processed. Amblyopia is characterised by a degree of loss of binocular function as manifest by poorer stereoacuity. It has been proposed that this is in part as a result of increased *inter-ocular* ***suppression*** one consequence of which is a reduction in normal binocular summation, the ability of the visual system to combine information across the eyes to detect a target. To obtain a direct experimental measure of suppression, psychophysicists can use ***dichoptic-masking*** (Legge, 1979), where a target stimulus (varied in incremental steps, e.g. of contrast or coherence) is presented to one eye and simultaneously an interfering mask (with the same properties as the

[5] Judgement of the direction of motion elicited by dichoptic presentation of two gratings presented 90 deg. out of spatial and temporal phase, and matched for perceived contrast.

[6] Multiple-choice stereoacuity test assessing fine depth perception. The test requires the observer to discriminate the circle that floats in front of two reference circles. Monocularly visible contours help to separate the form from random-dots background. Ten disparities are probed from 20-400sec of arc, at ~40cm viewing distance.

target, at a fixed stimulus-level) is presented to the other eye. Normally the mask elevates the stimulus-level required to detect the target. Huang et al. reported essentially normal masking in amblyopes, when the visibility of dichoptic stimuli was matched across the AE and FE, otherwise a weaker effect was observed for AE masking (P. C. Huang et al., 2012), suggesting intact binocular processing in this group. The non-dominant eye (AE in amblyopic observer) may be more affected by masking in the dominant-eye (FE) because of less interocular inhibition than normal.

Others have used supra-contrast-threshold global processing tasks - ostensibly motion and form discrimination of multiple elements to probe the conditions under which binocular combination can occur, showing that adding contrast energy to the AE could lead to balanced global integration e.g. global motion processing across the eyes (Ding & Levi, 2011; Narasimhan, Harrison, & Giaschi, 2012). Using a different paradigm Huang et al. presented dichoptically supra-threshold sine-wave gratings in different-phase and the perceived centre of the middle dark stripe of the cyclopean image was measured (C. B. Huang, Zhou et al., 2009; C. B. Huang, Zhou et al., 2011). The contrast required for the AE-image to contribute equally to the percept was substantially higher than for the FE attesting the presence of interocular imbalanced contribution to binocularity in amblyopia.

At present, there is not yet a unique definition (nor a measuring technique) for interocular suppression and possible specific mechanisms active in amblyopia are under investigation.

2.4 Physiological basis of amblyopia

2.4.1 The critical period

Ocular Dominance (OD) of individual V1 neurons has been characterised using a “7-point-scale”, from Hubel and Wiesel (1962). This measure quantifies the extent to which a cell could be driven by either or both eyes. In Figure 15 we plot the number of cells falling into each category: bar 1 represents all-contralateral-dominance cells, bar 7 all-ipsilateral-dominance and intermediate levels show the degree of eye preference in between these extremes, with bar 4 corresponding to cells with no eye-preference (i.e. equally well-driven by stimulation through either eye). Hubel and Wiesel showed that experimentally induced Monocular Deprivation of visual stimulation (MD through eye-suturing, eye-muscle surgery or the use of an opaque lens) dramatically changed the functional architecture of cortex if it is imposed during a “critical period” after birth, both

in kittens (Hubel & Wiesel, 1970b) and monkeys (Hubel, Wiesel, & LeVay, 1977)–see Figure 15.

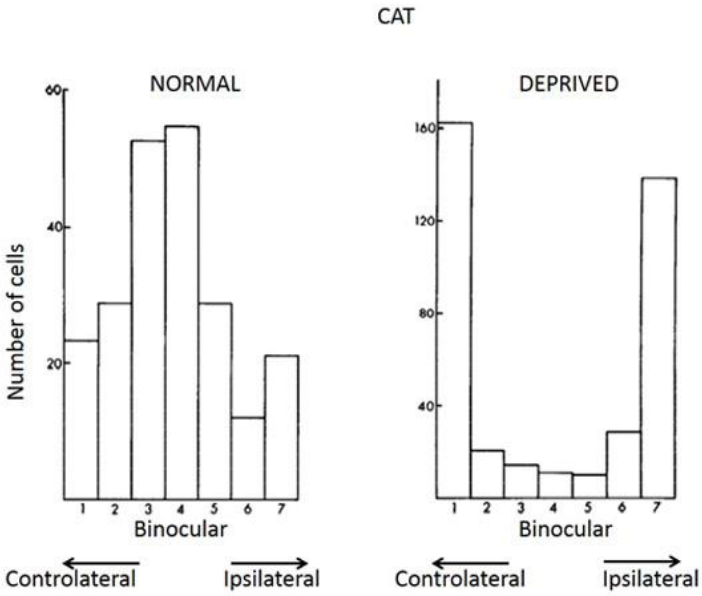


Figure 15 Histogram of Ocular Dominance for V1 neurons in a normal adult cat (left) and the average distribution of OD in four kittens raised with right MD induced –muscles resection- at 3 to 18 weeks after eye opening (right). Column 1 means 100% preference for the left eye; column 7 means 100% preference for the right eye. Normally, ~ 85% of cells does not have a strong preference for either eye (scoring 3-5), i.e. are binocular. Binocularly "balanced" cells (col4) are under-represented in MD kittens, as visible in the right graph. This highlights the susceptibility of visual cortex during a critical period (from birth to ≈6months in cats) for OD to shift away from the deprived eye (preference for left eye, col.1 compared to right eye, col.7) and associated reduction of binocular cells (adapted from Hubel & Wiesel, 1965a), p.1049 (related copyright licence agreement No 4414811341577)

Initially, the authors surgically sutured one eye of kittens (and later baby monkeys) and discovered that MD (of light and form) led to substantial shifts in ocular dominance towards an increased cortical representation of the unaffected eye. This, the authors argue, is consistent with OD resulting from binocular competition between monocular inputs. The notion is then that MD leads to reweighting of each eye's contribution to the final percept –since V1 neurons are more weakly driven by the deprived eye (Wiesel & Hubel, 1963). Furthermore the effects of MD were visible only *if* one eye was deprived *and* the other was not, i.e. suturing both eyes did not double effect size (Wiesel & Hubel, 1963). This means that the mechanism supporting shifts in OD was based on active competition and not disuse. The effect of MD was reversible - after a period of reverse occlusion – but only when treatment occurred during the *critical period* (Wiesel & Hubel, 1965a). The functional consequences of MD declined with age (until 3 months from birth) and were absent in adult cats (Wiesel & Hubel, 1965b). Their work highlighted the critical role of binocular visual experience vs. deprivation on the normal development of visual

system in animals (cats and monkeys) during a sensitive period. This is the period of maximal vulnerability to abnormal visual stimulation, called “critical period”.
Figure 16 reproduces possible effects of abnormal development compared to normal, in macaque monkeys.

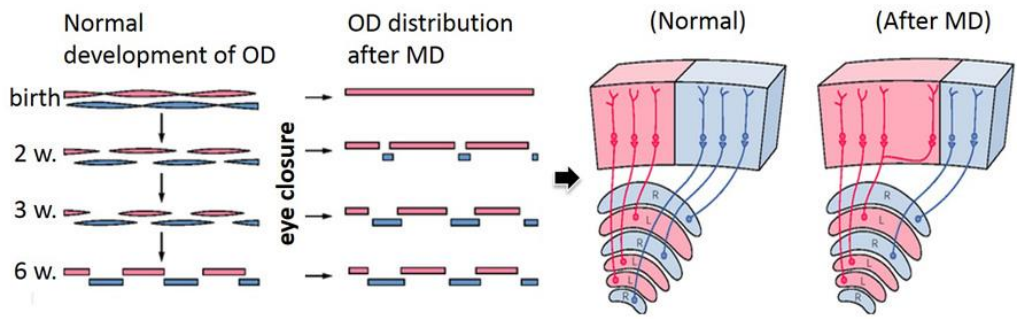


Figure 16 Normal and right-eye deprived development of OD **(from the left)** at birth (experience-independent) and during the following 2, 3 and 6 weeks (descending rows, respectively). Pink colour for left eye, blue for right eye; on the right of the figure (using the same colours for eyes' identification) consequent effects on anatomical structure and the distribution of physiological connections from LGN (below) to V1 (OD schematic representation just above) respectively in normal monkeys (left) and deprived animal (all right). The shrinkage of deprived eye's column is evident. Adapted from Hubel et al. (1977), p.404 (related copyright licence agreement No 4414811341577)

Different species show different sensitive periods to visual experience, including induced or naturally occurring MD. The most sensitive time-span is commonly referred as 4 to ~16 weeks after birth in kittens or postnatal to ~1 year in monkeys (Hubel & Wiesel, 1970b; Hubel, Wiesel, & LeVay, 1976; Hubel et al., 1977) and increases in babies, varying for each visual function (Daw, 1998) – see also section 1.4. In a subsequent phase of neural refinement, the tuning of cells (in particular for eye-preference –OD formation- and orientation-selectivity) is refined based on visual experience.

Hubel and Wiesel found that (i) a prolonged period of MD causes maximal loss of responsiveness of cortical neurons driven from the deprived eye (Figures Figure 15 and Figure 16), (ii) binocular deprivation does not affect OD formation and (iii) alternating MD and binocular activity decreased binocular cells' activity - (reviewed in Wiesel, 1982). The authors considered as *critical* the period during which deprivation was effective (in cats and monkeys): a few days of MD in the post-natal period was enough to reduce the proportion of the area of V1 driven by the deprived eye, so that afterwards only few cells responded to stimulation through this eye. Hubel and Wiesel proposed that activity-dependent competition (within layer 4 of V1) is the driver for OD column formation. Further areas in the visual pathway, where monocular inputs are integrated,

have a longer period of development and so a prolonged MD may generate long-term impairments, involving competition between each eye's image (Hubel & Wiesel, 1965b).

Importantly, the developmental stage when MD occurs depends on 1) the visual area under consideration 2) the visual property deprived (e.g. orientation, direction, contrast or depth) and 3) the species deprived (e.g. cats, monkey, mice or human) plus the individual rearing history. Thus, the critical period may correspond to the period of development of the specific ability (from birth) or to the period of possible recovery after MD (Daw, 1998).

Considering the consequence of depriving an animal of a specific *feature*, orientation-selectivity is vulnerable in kittens being exposed to only a single orientation (reared wearing head-goggles; Tanaka, Tani et al., 2009), mice (B. Wang, Sarnaik, & Cang, 2010), monkeys (using tetradotoxin, TTX drops vs. bilateral lid suture- Blakemore, Garey, and Vital-Durand (1978) and ferrets (Chapman & Stryker, 1993; White, Coppola, & Fitzpatrick, 2001). In all cases, a normal maturation of orientation-selectivity required neural activity driven by binocular visual experience to occur before completion of OD formation in V1-4 (Sengpiel & Kind, 2002). Other studies investigated direction-selectivity (DS) of V1 cells in kittens (Hubel & Wiesel, 1963), macaques (Hubel & Wiesel, 1968) and ferrets (Y. Li, Fitzpatrick, & White, 2006) – e.g. Li et al. reared the animal, for 3 days post-natal, in a 12 hrs light-dark cycle, while exposed to a series of wave gratings, differently modulated in direction of motion. Like orientation-maps, V1 has an ordered map of direction-selectivity with OD and orientation selectivity predicting DS. DS has a shorter CP: the effect of post-natal rearing history on DS appears irreversible, e.g. dark adapting for 3 weeks after eye-opening a ferret, also when restoration of normal visual experience occurs later-on (Y. Li et al., 2006). Finally, for a better definition *of "critical period"* and its complexity, we should refer to luminance, contrast sensitivity, frequency-tuning and stereopsis. An interesting study on monkeys highlighted the different duration of CPs relative to different visual properties: luminance (birth to 3months), contrast (birth to 6m), SF-tuning (birth to 18m), and stereopsis (birth to 24m) (Harwerth, Smith et al., 1986).

In summary, the critical/sensitive period extends from eye-opening to completion of visual system development and is greatly affected by MD and visual-feature(s) deprived (Harwerth et al., 1986). During this period eye-dominance is determined (DE-dominant eye; NDE-non-dominant eye) as well as functional-cortical-maps (selectivity for orientation, direction etc.). Cortical mechanisms supporting such neural refinements

include (for single neurons) response modification, dendritic pruning and terminal sprouting and (for populations of neurons) synaptic realignment and axon terminals retractions (Espinosa & Stryker, 2012). Neural circuits are shaped by visual experience to determine the complex pattern of connectivity, characteristic of the mature visual system.

In amblyopia, three '*critical* periods' (CPs) are distinguishable based on the magnitude of acuity-change in the AE (Marianne Piano, O'connor, & Stephenson, 2015)-see Figure 17. These periods are based on either the development of the visual system ('development CP'; discussed in section 1.4) or on the onset of amblyogenic factors ('sensitive CP': the later the onset the less disruptive the effect on visual acuity) or at the response to treatment that is likely to be obtained ('recovery CP'). The end of the recovery-CP is currently undefined (in Figure 17: the 'Recovery-CP' line never reaches zero) and it seems that amblyopia can be treated also in adults (see section 2.8).

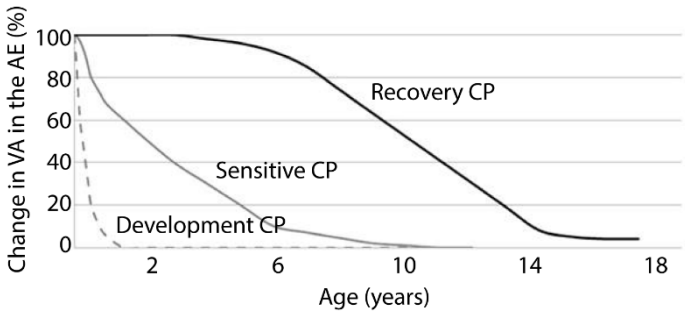


Figure 17 Schematic representation of three critical periods (CPs) in amblyopia. Developmental, sensitive and recovery CPs are defined based on the effect on monocular acuity in the AE: uncrowded visual acuity develops up to the age of 2-3 yrs and is most sensitive to amblyogenic factors until 3-5 yrs, less afterwards, until ~12 yrs; treatments might be effective up to adult-age, but are most effective up to 6-8 yrs. From M. Piano et al. (2015) – http://doi.org/10.22599/bioj.18: distributed under the terms of the CreativeCommons BY 4.0.

2.4.2 Binocular vision in amblyopia

In order to obtain a single percept from binocular stimulation either (a) binocular-summation occurs normally or (b) mechanisms operate to reduce interocular competition of discordant monocular representations. When identical images are presented to the two eyes the final percept is a fused image but when two *sufficiently different* images are presented to the two eyes, our perception can alternate continuously over time between the two incongruent inputs. Binocular perceptual phenomena have been described in section 1.3.

Wheatstone argued that ***binocular rivalry***, alternation of percepts, was a special case when fusion breaks down. Today, researchers still debate whether the prolonged

perception of the stronger stimulus when rivalry images are imbalanced is caused by preferential access to OD (excitatory mechanism) or, more probably, to reduced suppression of its signal (inhibitory mechanism; Blake & Logothetis, 2002). ***Interocular suppression****, a perceptual and cortical mechanism,* can arise when the quality of monocular images is constantly dissimilar, as in cases of constant unilateral blur in presence of anisometropia or constant offset in retinal images caused by strabismus. This cortical phenomenon is activated by temporal and spatial properties of competing stimuli (P. C. Huang et al., 2012) and is thought to avoid diplopia through "active cortical inhibition of objects in all part of the visual field of one eye" (Jampolsky, 1955).

If a prolonged and constant MD during the critical period occurs, one view is that the resultant suppression **causes** amblyopia. This hypothesis has been investigated as active cortical mechanism mediated by inhibitory connections (Blake & Logothetis, 2002). Intuitively, suppression may thus explain amblyopia in presence of strabismus, driven by the misalignment between eyes (Sengpiel, Jirmann et al., 2006): the deviated eye-AE- image is suppressed. In term of mechanism, fewer excitatory intrinsic horizontal connections between neurons from the deviated eye would be active (Harrad, Sengpiel, & Blakemore, 1996). A similar phenomenon occurs for anisometropic amblyopia, where suppression is (somewhat less effectively) driven by defocus in the AE: higher levels of suppression are *associated* with greater severity of amblyopia (higher refractive error), as larger IOD in acuity and poorer stereopsis (Lai, Alexander et al., 2011). Less is known about the physiological basis of suppression in anisometropia, although there is evidence for a different suppression-mechanism occurring in humans, since anisometropes show a different pattern of functional deficits to strabismics (Narasimhan et al., 2012).

Independent of the kind of amblyopia, IOD in acuity is correlated with the amount of interocular suppression (J. Li, Thompson et al., 2011) and the extent of central suppression (within 20 degrees of visual angle) (Babu, Clavagnier et al., 2013). This suggests a primary-causative role of suppression in the mechanism of amblyopia. Below we discuss the possibility that suppression may not be causative but an adaptive response to avoid diplopia.

2.4.3 Cortical locus: Downstream mechanisms

Since the pioneering studies of Hubel and Wiesel, V1 has been thought to be the neurological basis for amblyopia. OD and the effects of MD are consistent with structural modifications of V1 modifying functional vision in amblyopia (see also section 2.4). The

excitatory binocular connections in V1 are likely to be reduced after degradation of the monocular input through the AE. Over the years a higher importance has been attributed to cortical regions linked to binocular vision (see section 2.4.3). Kiorpes and colleagues compared psychophysical measures and physiological recordings in V1 in macaques (Kiorpes, Kiper et al., 1998). They measured spatial resolution and contrast sensitivity in monkeys reared with MD (3-defocus; 3-esotropia) and found that the magnitude of behavioural deficit was greater than predicted by the physiology, and was dependent on the severity more than type of deprivation. Behaviourally, animals could not detect higher SFs through the AE, in the presence of (albeit reduced) cortical activity in V1 in response to such stimuli. This suggests V1 is not the only cortical area limiting visual function in amblyopia.

More recently, it has been suggested that moderate differences in the activity of V1 single-cells arising from amblyopia might be amplified in the pooled activity of cells in downstream areas. To test this Shooner, Hallum et al. (2015) have measured behaviour and recorded from large populations V1 and V2 cells, in macaque monkeys reared with blur, strabismus or no intervention. Monocular contrast sensitivity was assessed using a 2AFC task to compute eye-dominance profile (60 samples of response for each electrode to define orientation, SF and eye preference, typical of the isolated OD). After 2 to 17 years, the physiological response to noise-SF filtered patches was recorded. Physiological responses correlated with changes in the CSF: greater difference in contrast sensitivity across the eyes correlated with strength of FE-preference in neural-responses. This was the case in V1 and V2 and was most evident for centrally presented stimuli. Such result suggests that vision through the AE might be limited by a weaker response at the single-cell level in V1, and by a reduction in decoding efficiency in downstream areas (V2 and beyond; linear discriminant analysis of responses pattern from a given neural population – N electrodes, each one averaged on 60 samples). Thus, suboptimal pooling beyond V1 might be the amblyopia mechanism (Shooner et al., 2015).

2.5 The role of cortical plasticity after abnormal visual development

Compensatory changes in OD following MD are an expression of **cortical plasticity,** i.e. the ability of neural circuits to remodel based on experience - for review: Polat and Sagi (1995). Case studies show reduced activation of cells driven by the deprived eye in both ipsi- and contralateral LGN that could result from local influences of eye

dominance or from imbalanced feedback from V1 (Barnes, Li et al., 2010; Blakemore & Vital-Durand, 1986), indicating anomalies in the feedforward and feedback connectivity between LGN and visual cortex (X. Li, Mullen et al., 2011). Although cortico-geniculate feedback may play a role, the neural basis for amblyopia seems to be cortical, as indicated, for example, by the significant shift in ocular dominance in V1 resulting from monocular deprivation. Experimentally induced strabismus in kittens of 8-10 days old produced a strong aggregation of cells according to eye dominance, shifting towards the non-deprived eye: ~60% of single cell recordings were driven by the non-deprived eye (Hubel & Wiesel, 1965a). Following the "7-point-scale", it seemed that cells falling in groups 2 and 3 in a normal cat, shifted to column 1 (i.e. preferentially driven by the contralateral eye) and groups 5-6 shifted, more moderately, to column 7, leaving only ~20% of cells being well driven by binocular stimulation. The same results are observed after alternated occlusion (opaque contact occluder on one eye every-other-day for ~10weeks), when only 9% of cells maintained a binocular preference (Hubel & Wiesel, 1965a).

Adult cats that have strabismus surgery (severing the rectus muscle) before the critical developmental period exhibit increased inhibition of synaptic connections driven by the deprived-eye (in vivo cells recordings; in Scholl, Tan, & Priebe, 2013). In term of mechanism, there is an effective suppression of activity deriving from the AE (specifically, reduced evoked potentials compared to normal animals). The authors argued that higher inhibition (here, the IOS mechanism) has a functional role in reducing diplopia. In general, simple cells become more monocular (with more being driven exclusively by the non-deprived eye) and complex cells become less tuned for disparity (Scholl et al., 2013). In term of anisometropia, an inter-ocular difference in blur creates *imbalanced* stimulation of V1 receptive fields, which in turn leads to more disorganisation of ocular dominance, with a consequent disruption of projections to later visual areas (especially V2). Such disrupted projections could lead to perceptual distortions of space without inducing inter-ocular suppression (Tao, Zhang et al., 2014). However, this distinction between plasticity mechanisms in the two types of amblyopia remains a hypothesis.

With respect to MD, a period of "reverse occlusion" (opening the previously-sutured eye and suturing the previously-open eye) can reverse changes in OD. Such cortical plasticity suggests that recovery from amblyopia should be possible by manipulating the animal's **visual environment** (Hubel et al., 1977). Inducing anisometropia using artificial blur (induced by means of a translucent-opaque contact lens) leads to a

selective rearrangement of effective input from the deprived eye as a result of a weakening of the excitatory signal from the affected eye (Movshon, Eggers et al., 1987). This leads to imbalanced activation of binocular neurons (Kiorpes et al., 1998; Movshon et al., 1987).

Mitchell examined the effect of a daily period of binocular exposure alternated with monocular patching on development of acuity in kittens that were raised in complete **darkness** for half a day (from 4 to 8 weeks of age). Therapeutic results improved when 30 minutes of daily binocular stimulation were alternated to up to 10 hrs patching (with 2 continuous hrs, AE and FE acuity were similar) and the risk of recurrence after treatment cessation reduced (Mitchell, 2008). A concordant input from both eyes would indeed have a leading impact in shaping cortical connections during development.

There is evidence that formation AND transmission of a concordant and **balanced binocular visual** input plays a fundamental role in the development of a normal visual system.

2.6 Pharmacological studies

The advent of modern imaging techniques has allowed for investigation of the requirement for normal neural activity in cortex to develop binocularity during the 'critical' phase of visual system development. The blockage of inhibition with GABAergic modulation (using Bicuculline, a GABAa antagonist) interferes with an inhibitory intracortical circuit involved in OD spacing definition and prevents normal development of visual functional architecture (Hensch, 2005; White & Fitzpatrick, 2007)

It has been reported that **dark-exposure** (darkroom rearing) after MD can lead to recovery of spatial resolution (grating acuity; in kittens; Duffy & Mitchell, 2013). Darkness (typically around 10 days of dark-adaptation) can have different effects depending on when it is applied. If applied immediately (1 week after MD), it seems to allow a slow and parallel gain in both eyes, whereas if applied after 5 to 8 weeks post MD, it leads to a rapid gain in AE to match FE functionality (Duffy & Mitchell, 2013). In addition a modest enhancement of cortical plasticity has been shown in juvenile mice, dark-adapted for a week, following 4 weeks of MD (Erchova, Vasalauskaite et al., 2017). The proposed mechanism for changes induced by dark-exposure is that this condition leads to secretion of neurofilament protein or neurotransmitters (e.g. GABA) that can "freeze" brain development in the pre-dark-exposure state. This minimises the effect of

deprivation, maintaining a higher plasticity, typical of a younger brain (Hensch, 2005; Takesian & Hensch, 2013), thus reinstating plasticity beyond what would otherwise be the CP for recovery from MD (Erchova et al., 2017). Possibly, modifying such *functional brakes* is an effective therapy for amblyopia that can restore visual function into adulthood, in animals (Mitchell & Duffy, 2014) and possibly in humans (Bavelier, Levi et al., 2010).

Such approaches assume that V1 is not only the initial stage where binocular input is formed but also the first stage of visual processing where neural circuits and the activity of populations of neurons can be critically modified.

2.7 Neuroimaging studies

A recent review of animal electrophysiology and human imaging studies on amblyopia concludes that there is sound evidence for functional modification of neural connectivity within and beyond V1, in parieto-occipital and temporal cortices, regions that are particularly involved in binocular vision (Joly & Franko, 2014).

Although human neuroimaging has demonstrated structural abnormalities in LGN of individuals with amblyopia (Barnes et al., 2010), the majority of studies only reported reduced activity in area V1 (Demer, von Noorden et al., 1988; Hubel, 1982; Sengpiel et al., 2006). More recently, physiological (Bi, Zhang et al., 2011; Sincich, Jocson, & Horton, 2012) and imaging studies (Barnes, Hess et al., 2001; Conner, Odom et al., 2007; Demer, Grafton et al., 1997; X. Li, Dumoulin et al., 2007) have focused on the development in amblyopic subjects of long-range synaptic projections from V1 to area V2, and also between downstream areas (Ho & Giaschi, 2009; Secen, Culham et al., 2011). White-matter pathways within the visual system have also been recently investigated: adults with amblyopia showed increased mean diffusivity in thalamo-cortical (but not cortico-cortical) pathways (Allen, Spiegel et al., 2015; Duan, Norcia et al., 2015). In a recent fMRI study, the interhemispheric functional connectivity reported under resting-state in patients with either anisometropic or strabismic amblyopia was altered compared to that reported in non-amblyopic viewers (Liang, Xie et al., 2017). These anomalies also revealed further insight into the abnormal connectivity characteristic of each sub-type of amblyopia.

Farivar, Zhou et al. (2017) measured the difference between fMRI BOLD response in V1, V2 and V3 to visual stimulation of the AE and FE (using a multifocal fMRI paradigm)

of anisometropic amblyopes. They report both attenuated BOLD response and distorted retinotopic maps in response to stimulation of the AE. Attenuated cortical activity could account for loss in sensitivity to visual contrast. Distorted retinotopic mapping could explain the perceptual spatial distortions that patients frequently report when viewing through the AE. The distortions of retinotopic mapping would result from a positional disorganisation of visual receptive fields (*disarray*) representing the AE input. In fact, in the central visual field, by analysing fMRI data using a population receptive fields (pRF) modelling, the enlarged pRF sizes for the AE could reflect an increased dislocation (/disarray) of the position of pRFs representing the AE or a reduction/loss of neural resolution (Clavagnier, Dumoulin, & Hess, 2015). Either processes would justify the reduced visual acuity in amblyopia. Interest in mapping deficit (focusing on inter-ocular differences, more than solely on the AE signal disorganisation) is increasing, as a viable aspect in future amblyopia therapies.

2.8 Treatment for amblyopia

2.8.1 Current clinical practice (occlusion therapy)

Current interventions for *childhood* amblyopia consist of a monocular treatment, to promote the use of AE before the age of 7-8 years, thought to be the upper limit of "critical period" to recover amblyopia (Ciuffreda et al., 1991). However, treatment should be offered to older children and teenagers as well: treatment success is now thought to depend on age *and* also baseline acuity, onset of amblyopia, history and compliance to previous amblyopia-treatment and concomitant health conditions (David K. Wallace et al., 2018). The general purpose of existing monocular strategies is to improve visual acuity, being the same level of vision in both eyes the best possible outcome to achieve. This goal in mind, the preferred practice to treat amblyopia consists on an initial period of "refractive adaptation" (RA; wearing of prescribed optical correction for 12 to 24 weeks) – also referred to as "**optical treatment**" (OT); then, should the interocular difference at BCVA persist, the FE is penalized, either by patching or - less frequently - using eye-drops (Stewart et al., 2011).

A period of adaptation to glasses alone demonstrated possible recovery (IOAD≤0.1 logMAR) in 27% of anisometropic children (Cotter et al., 2006; Stewart et al., 2011) and 32% of children with strabismic and combined amblyopia (Cotter et al., 2012). The relevance of optical treatment of amblyopia deserves attention also in older children and adults, as preliminary results reported a significant improvement (or stabilised acuity), following 4 to 16 weeks of wearing corrective lenses, in ~20% of cases (T.Y. Gao,

Anstice et al., 2018). The logic of penalization therapies is to overcome FE dominance to force the use of AE to improve functional vision. Penalization (or occlusion) methods include pharmacological treatment and patching (current amblyopia preferred practice pattern in David K. Wallace et al., 2018). The first consists of eye-drops, usually 1% atropine in ophthalmic solution, applied in the FE to produce cycloplegia (prolonged pupil dilation) for ~3 days. In this way, close-up vision through the FE results blurry, thus encouraging the alternative use of the AE (i.e. distant vision will be non-directly affected). Instead, patching consists of stopping vision through the FE by applying an opaque patch (usually, an adhesive patch directly on the skin; when necessary, glasses are worn on top), thus promoting the use of the AE. To avoid skin irritation, the patch may be applied over the eyeglasses, but the child might find ways to look around it, thus causing the treatment to be interrupted. From children and parents' perspective, atropine seems to have a less negative impact than patching on treatment compliance and social stigma (Felius, Chandler et al., 2010), although a transient reduction of visual acuity in the FE has been reported more frequently in relation to atropine than not patching treatment (Pediatric Disease Investigator Group., 2003).

Treatment duration and dose depend on individual response; estimates of current treatment-durations vary between 3-6 months (Repka et al., 2003; Stewart et al., 2007; Taylor & Elliott, 2014) up to 1-2 years (Awan, Proudlock et al., 2010), typically involving 2-6 hours patching per day (Stewart et al., 2011). **Occlusion** does lead to significant improvement in visual acuity – at least 0.2 logMAR - in ~70% of cases (Stewart et al., 2011), with similar results after weekend atropine or daily patching (Pediatric Eye Disease Investigator Group., 2004), possibly maintained also at 15 years after treatment completion (Repka, Kraker et al., 2014). The use of oral levodopa as a potential adjunction to patching in reducing residual amblyopia has been found not effective (Repka, Kraker et al., 2015). Thus, the need to find alternative ways to solve residual amblyopia, resulting in ~1/3 children, who show an incomplete response to occlusion (Scheiman, Hertle et al., 2005). Moreover, the impact of such treatments on binocular visual function is less certain (D. K. Wallace, Lazar et al., 2011) in that stereopsis does not improve in more than 50% cases (Stewart et al., 2013). Stereopsis is particularly un-amenable to treatment in strabismic amblyopia (McKee et al., 2003). This may be associated with active suppression of the AE (Sengpiel et al., 2006) and abnormal neural responses to stimuli driven by AE, as discussed in section 2.4.

These traditional treatments have been found to be associated with a relatively high risk of recurrence, of around 30% one year after treatment cessation (Bhola et al., 2006; J.

Holmes, Beck et al., 2004). A two-year-follow-up study demonstrated a reduction of AE acuity of at least 0.2 logMAR in 20% of patients (Repka, Wallace et al., 2005), although a more recent report found no recurrence even up to 10-15 years following treatment (Repka et al., 2014). Another critical limitation of penalization therapies is that adherence is poor (around 45% in Pradeep, Proudlock et al., 2014) and so is the elicited compliance (i.e. received versus prescribed treatment-dose), estimated at 44% (M. P. Wallace et al., 2013) and confirmed at 33-58% in monitored-patching studies (for a review: Stewart, Moseley et al., 2017). This figure is even lower in older children (J. M. Holmes, Lazar et al., 2011). In short, difficulties in implementing the treatment reported by families and the negative psycho-social impact of wearing a patch limit compliance and prevent successful treatment (Carlton & Kaltenthaler, 2011; Loudon, Passchier et al., 2009).

2.8.2 Monocular alternative treatments (new-occlusion techniques and PL)

Instead of continuous patching, *intermittent-occlusion* consists in occluding the FE for e.g. ~66% of the glasses-wear time: a thin glass-layer of liquid-crystals is coupled to the corrective lens, blocking vision e.g. for ~40s every minute (Spierer, Raz et al., 2010). Crystals are oriented applying a voltage, controlled via a pre-programmed microchip, connected to a coin-battery. This technique seems to be equally effective as traditional patching (J. Wang, Neely et al., 2016), but requires wearing glasses for a longer amount of time every day, which might be a disadvantage to secure high compliance (limiting the therapeutic benefit). Alternatively, but only in case of mild amblyopia, a *translucent filter* (Bangerter; Ryser Optik AG, St. Gallen, Switzerland) applied to the FE corrective lens can be equally effective as a fully-opaque patch (Rutstein, Quinn et al., 2010).

Across traditional and alternative occlusion therapies, the integration to glasses of an *occlusion-dose-monitor* device is highly recommended to record the individual dose-response, to incentivise adherence to treatment and, eventually, to personalise treatment-plan (Stewart et al., 2017). Although dose-monitor devices help achieving higher levels of compliance, that compliance remains usually poor limits the positive therapeutic outcome of any modality of patching-therapy.

Using the Cambridge stimulator (known as CAM) a new approach to treating amblyopia was introduced, that focused on practicing close activities while wearing a patch (on the FE). The preliminary study (Campbell, Hess et al., 1978) showed a positive effect on acuity of repeated sessions (on average 7, 7min long) during which the child played

drawing games, on a Perspex plate, while forced to focus on a series of rotating discs, beneath the plate, each representing a high-contrast grating at a certain spatial frequency. Although the treatment failed to demonstrate its efficacy in a following trial (Tytla & Labow-Daily, 1981), the concept of treating amblyopia by *stimulating* the AE instead of just depriving the FE raised a strong interest among researches.

Developing this idea, more recent alternatives to occlusion therapies have targeted binocular function. With the aid of repetitive perceptual training at psychophysical tasks (such as vernier or grating acuity), patching may improve monocular and binocular abilities. **Perceptual learning** (PL) paradigms applied to the treatment of amblyopia have focused on contrast detection (Polat, Ma-Naim et al., 2004), crowded acuity (Z. Hussain, Webb et al., 2012) and stereo-depth discrimination (Xi, Jia et al., 2014). Generalizability to untrained tasks may occur (Zhang, Cong et al., 2014), but is poorly understood. For example, contrast training transfers to acuity but not vice versa (A. T. Astle, Webb, & McGraw, 2011). PL training while applying “inverse occlusion” (i.e. patching the AE) was also effective (Agervi, Kugelberg et al., 2009; J. Zhou, Reynaud, & Hess, 2014a; J. Zhou, Thompson, & Hess, 2013). However, these therapies are highly time-consuming, highly repetitive and therefore likely to engage low compliance (particularly in children).

2.8.3 Binocular alternative treatments (and brain-stimulation)

With a duration much shorter than patching (1 week to ~2 months) binocular therapies still produce comparable gains in acuity to patching but might also improve stereoacuity, having a functional impact, e.g. in controlling eye-hand coordination, on everyday life (O'Connor, Birch et al., 2010). Until recently, two main alternative binocular treatments that combine PL-training with dichoptic presentation, have been proposed by other groups of research: (1) “**Anti-Suppression Therapy**” and (2) “**Interactive-Binocular Treatments**”. Hess and colleagues have developed a binocular “Anti-suppression” therapy to reduce the suppressive role of FE over AE’s signal. Here, dichoptic content is presented in the context of a game that can only be played using both elements of a binocular image-pair. To allow for dichoptic view, they used either (i) lenticular overlay and red-green glasses (E. E. Birch, Li et al., 2015; R. F. Hess, Mansouri, & Thompson, 2010a, 2010b; S. L. Li, Jost et al., 2014) or (ii) a head mounted display (Black, Hess et al., 2012; Knox, Simmers et al., 2012) or (iii) lenticular overlay applied to a full-sized screen (To, Thompson et al., 2011). Tetris is the most commonly used game: AE receives the falling block that needs to fit the bottom puzzle presented across eyes. Monocular signals, initially favouring AE, are progressively balanced in contrast to

overcome inter-ocular suppression. Visual acuity and stereopsis improved (around -0.2 logMAR and +200arc sec in >60%) after 1-2 hours per day for 2-6 weeks on treatment (R. F. Hess et al., 2010a, 2010b; R. F. Hess, Thompson et al., 2012; To et al., 2011). These methods have been successfully tested both in adults (Black et al., 2012; R. F. Hess, Mansouri, & Thompson, 2011; R. F. Hess et al., 2012; To et al., 2011) and children (Knox et al., 2012; S. L. Li et al., 2014). In adults, it seems that *binocular* gameplay leads to greater gains compared to a monocular version (J. Li, Thompson et al., 2013) and monocular-movie viewing (Vedamurthy, Nahum, Huang et al., 2015). Dichoptic video-game play also seems to improve disparity perception (Xi et al., 2014). Such improvements have also been demonstrated in children (S. L. Li et al., 2014). However, a 16 weeks randomised-controlled-trial in children 5 to <13 years old reports higher improvement in the AE vision after patching therapy (2 hrs a day; mean VA gain in the AE: 0.11 logMAR) than after binocular-gameplay (1 hr a day of Tetris game; mean VA gain in the AE: 0.14 logMAR) (NCT02200211; J. M. Holmes, Manh et al., 2016). The reported adherence to gameplay was limited in either treatment-group (only about a fifth exceeded 75% adherence) and ~80% received previous treatment for amblyopia (other than glasses). Conversely, only 2 weeks of training (on a smaller sample, of same age children) produced contradictory results: dichoptic gameplay was more effective than 2 hours patching (K. R. Kelly, Jost et al., 2016). A newly published randomised controlled trial on older children, teenagers and adults (ACTRN12613001004752; T. Y. Gao, Guo et al., 2018) compared the effectiveness of dichoptic modified Tetris game (falling blocks visible to the AE only) versus a non-modified version of the same game: interestingly, both groups showed a gain in acuity in the AE (respectively, mean gain of 0.06±0.12 and 0.07±0.10 logMAR) after 6 weeks (dose prescribed: 1 hr per day). Therefore, it has yet to be established whether binocular gameplay is more effective than patching, in children, and if, at any age, the effectiveness of this treatment is due to the nature of visual stimulation or to the high compliance elicited by game-based therapies.

Direct brain-stimulation (rTMS or tDCS) for a limited period (~a week) can accelerate the effect of therapeutic training on adults (R. F. Hess & Thompson, 2013; Spiegel, Byblow et al., 2013; Spiegel, Li et al., 2013). Training with dichoptic gameplay displayed on a virtual reality (VR)-head mounted device (where some details of the scene are shown only to the AE, while the main ‘acting-character’ is shown only to the FE) elicit significant improvements in vision (mean gain: 0.15 logMAR) in amblyopic adults after only 8 sessions, of 40 minutes each (Ziak, Holm et al., 2017).

Herbison and colleagues have developed a system based on video footages and video-games 0.(moving on, from an initial virtual reality setting; Eastgate, Griffiths et al., 2005) to train both eyes by mean of shutter glasses. This technology allows a dichoptic presentation or visually related images with some details visible only though AE's lens while the patient plays an interactive game. Even if stereopsis was not stimulated, ~2logMAR lines gain in visual acuity was obtained in 4-8 years old children after ~15-30 minutes, per one week (Cleary, Moody et al., 2009; Waddingham, Butler et al., 2005) or ~6 weeks (N. Herbison, Cobb et al., 2013). This therapy has gone through further investigation in a RCT study (NCT01702727; Nicola Herbison, MacKeith et al., 2016). The authors compared three dichoptic viewing methods for amblyopia treatment in children aged 4-8 years: video-footage viewing (video-images to the AE, while the same background was visible through both shutter lenses), modified game (targets to the AE only) and placebo/sham game (same view through either lens). Acuity in the AE moderately improved in all three groups (mean gain: 0.07 logMAR), showing both no substantial difference between arms and similar results to anti-suppression therapies.

Our alternative treatment proposal

In chapter 3, we describe a novel home-based BBV ("Balanced Binocular Viewing") treatment, which requires children to watch dichoptic movie stimuli for one hour per day, for between 2 and 6 months (depending on the outcome of regular orthoptic assessment). Dichoptic movies are balanced in visibility across eyes, by applying a level of Gaussian blur to the FE-image that is sufficient to match acuity to the AE. Stimuli are also modulated in horizontal disparity to stimulate stereopsis. Exploratory results from a group of 22 children (6-11 years) with anisometropic (7), strabismic (6) and combine mechanism amblyopia (9) showed a positive therapeutic effect of BBV at logMAR acuity (mean gain in the AE was 0.27 logMAR, σ=0.22 logMAR). Contrary to the notion that intra-ocular suppression (IOS) may have a causative role in amblyopia, IOS did not change significantly following treatment. Data on compliance, obtained via a simple task interleaved to movies, showed a high participation to BBV (on average, 89% of prescribed daily dose on 68% of days when the system was installed at home). Although on a small sample, these exploratory results brought evidence on the therapeutic effect and applicability of this easily adaptable and unsupervised therapy for amblyopia.

This chapter is adapted from the published paper (cited elsewhere in this thesis): Bossi, M., et al. (2017). "Binocular Therapy for Childhood Amblyopia Improves Vision Without Breaking Interocular Suppression." Invest Ophthalmol Vis Sci 58(7): 3031-3043; permission to reproduce the paper's material has been granted by its Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License.

3 Balanced Binocular Viewing treatment for amblyopia

Having considered the limits of occlusion therapies and a lack of consistency in the results obtained across different studies investigating the alternatives, we wanted to contribute to the developing awareness of the amblyopic syndrome and to the improvement of its management and treatment. We considered the treatment history of our potential participants prior to enrolment, to be able to isolate the specific effects of our proposed (binocular) treatment for amblyopia. We developed an engaging home-based treatment method, with the potential to elicit high levels of compliance and explored its safety and efficacy in improving vision of children with amblyopia. In parallel, by designing a new test, we aimed at exploring how binocular therapies affect the amblyopic visual function and how they work.

3.1 Introduction

We have dedicated Chapter 2 to a complete introduction of amblyopia (definition, associated functional deficit, physiological basis and treatment). To sum up, amblyopia is a developmental disorder of vision with a prevalence of 2-5% (Attebo et al., 1998), defined as a monocular (rarely binocular) reduction of the best-corrected visual acuity (BCVA, or simply 'acuity') in an otherwise healthy eye. Amblyopia is caused by a prolonged period of abnormal retinal stimulation (mainly) due to *strabismus* (ocular misalignment), *anisometropia* (refractive imbalance), or both *(combined)* and leads to functional deficits including reduced contrast sensitivity (Levi & Harwerth, 1977), poor spatial localization (Levi & Klein, 1983), poor stereovision (McKee et al., 2003) and foveal crowding (Levi & Klein, 1985).

Typically, amblyopia is treated only if the inter-ocular acuity difference between the amblyopic eye (AE) and the fellow eye (FE) is at least 0.2 logMAR (Powell & Hatt, 2009). Current treatment commences with 12-24 weeks of wearing prescribed optical correction, which improves AE acuity to normal levels in 27-32% of cases (Cotter et al., 2006; Cotter et al., 2012). Otherwise treatment to promote the use of the AE is administered, which consists of patching the FE (2-12 hrs/day; Stewart et al., 2011) or blurring the FE with atropine eye drops (T. Li & Shotton, 2009) for up to 24 months (Repka et al., 2005; V. Tailor, Bossi et al., 2016). Such occlusion therapies improve acuity in ~70% of patients by ~0.2 logMAR or more (Stewart et al., 2011). However, their impact on binocular vision is less certain (D. K. Wallace et al., 2011) and amblyopia recurs within a year in ~25% of <8yr olds (Bhola et al., 2006; J. Holmes et al., 2004). Moreover, compliance is poor: on average, only 44% of the prescribed daily dose is received in 58% of days ascribed for treatment (M. P. Wallace et al., 2013).

Central to current treatment is the idea of a *critical period* for visual development (see section 2.4.1). In humans, acuity and contrast sensitivity are adversely affected by periods of *monocular deprivation* respectively before the age of 10-12 years, even though adult-like performance is reached before, at 6 years (Lewis & Maurer, 2005). However, the notion that amblyopia is not treatable outside of this period, has been challenged by studies finding that adults 'forced' to use their AE show substantial improvements in contrast sensitivity (Polat et al., 2004), crowded acuity (Z. Hussain et al., 2012) and stereopsis (Xi et al., 2014).

Interocular suppression (IOS, or simply '*suppression*') is often considered to be central to the mechanisms underlying amblyopia, though functional definitions vary. When measured with a dichoptic motion-coherence task (Black, Thompson et al., 2011), suppression has been quantified as the contrast offset between the eyes at which binocular integration fails (Black et al., 2012; Narasimhan et al., 2012). Others have measured suppression as the 'effective contrast ratio' necessary to perform a dichoptic phase-alignment task (Kwon, Lu et al., 2014). Recently, some of us have developed a test measuring the contrast mixture of dichoptically-presented letter-pairs that leads observers to switch from using one eye to the other (Kwon et al., 2015). In terms of physiological mechanism (see also sections 2.4 and 2.5), animal models have shown that prolonged monocular deprivation leads to weakened excitatory drive from the deprived eye and hence imbalanced activation of binocular cortical neurons (Kiorpes et al., 1998; Movshon et al., 1987; Sengpiel & Blakemore, 1996; Wiesel & Hubel, 1963).

The resulting suppression is thought to be the result of an active inhibitory cortical mechanism (Sengpiel, Baddeley et al., 1998).

Stronger suppression is associated with more severe amblyopia (J. Li et al., 2011; Narasimhan et al., 2012) and has traditionally been viewed as an adaptive mechanism (to avoid *confusion*, between non-corresponding inputs; Holopigian, Blake, & Greenwald, 1986). In contrast, it has been proposed that suppression has a *causative* role in amblyopia, making it a candidate target for therapy (J. Li et al., 2011). Notably (see also section 2.8.2), Hess and colleagues have developed *"anti-suppression" therapy* which uses games (with elements split across the eyes to promote binocularity) to treat amblyopia in adults (R. F. Hess et al., 2010a; R. F. Hess et al., 2012; To et al., 2011) and children (E. E. Birch et al., 2015; Knox et al., 2012; S. L. Li et al., 2014). After 2-6 weeks of treatment, visual acuity improves by an average of ~0.15 logMAR and stereopsis was measurable in ~ 45% of participants (for the first time in ~2/3 of them). However, a randomised controlled trial of an alternative binocular therapy (iBIT) reports only modest success with children (mean acuity gain: 0.08 LogMAR; Nicola Herbison et al., 2016). Modern 'Perceptual Learning' treatments have also yielded positive results in adults and older children via monocular training (using the AE) on psychophysical tasks (Levi & Li, 2009; Polat et al., 2004) or video-game play (Bavelier et al., 2010; Zahra Hussain, Astle et al., 2014), as have hybrid approaches that interleave monocular task with dichoptic video-game play (Vedamurthy, Nahum, Huang, et al., 2015). A review of the literature on monocular and binocular behavioural-training methodologies can be found in Tsirlin, Colpa et al. (2015), or see section 2.8 for more on the treatment of amblyopia.

Although the changes in acuity and binocularity elicited by such therapies are widely cited as examples of cortical plasticity (Bavelier et al., 2010; Mitchell & Duffy, 2014) effected through a change in suppression (J. Li et al., 2011), in reality the mechanism(s) remains poorly understood. In this chapter, we extend what we introduced in the very last paragraph of section 2.8: we describe a new variant on alternative binocular therapies, namely Balanced Binocular Viewing treatment (BBV), which utilises dichoptic movies that are matched in visibility across the eyes. By relying on periodic vision assessments in clinic, we explored whether BBV was an engaging treatment, as we hypothesised, and (therefore) if it was effectively improving visual function -primarily acuity- in a relatively short treatment-period. The periodical clinical-checks and a remotely-controlled task, interleaved to movies, allowed us also to explore the safety (and not only efficacy) of BBV treatment.

Unfortunately, in terms of efficacy, we did not have an external control group, being limited in the resources. This is in line with the *exploratory* classification for this study, although it limits the generalisability of our results.

In short, our exploratory study investigated a procedure that was designed to be both an effective and safe home-based treatment, engaging a high level of compliance, and a platform for exploring how binocular therapies work.

3.2 Methods

3.2.1 Participants

3.2.1.1 Potential participants

A total of 40 children (3 to 12 yrs; 17 males) were identified as potential participants from the Richard Desmond Children's Eye Centre (RDCEC) at Moorfields Eye Hospital or its partnered local services (London): they all showed a reduced acuity despite the absence of other ocular pathologies. Specifically, each child was potentially eligible if showing (a) a reduced acuity in the AE (>0.2 logMAR) *and* (b) at least 0.2 logMAR inter-ocular difference in the best-corrected acuity. Before being eventually recruited, each child had to complete a period of *Optical Treatment* (OT), consisting of wearing the up-to-date prescription, full-time, for a minimum period of 16 weeks.

Inclusion and exclusion criteria. All children received OT prior to participating in our study. Following this period, the first 8 children without-strabismus and 16 with strabismus who gave their assent (and whose parents provided consent) were included if **(i)** acuity in the AE did not improve on two consecutive visits (#1 and #2; eight weeks apart) and **(ii)** a poor vision persisted, i.e. if their amblyopia was confirmed.

Amblyopia was defined by IOAD≥0.2 logMAR *and* FE acuity≤0.2 logMAR. We considered anisometropic, strabismic or combined mechanisms amblyopia. Specifically, amblyopia was classified as *anisometropic* if there was a difference of at least 1 dioptre (D) in spherical equivalent or 1.5D in astigmatism between the two eyes. *Combined* mechanism amblyopia was also associated with heterotropia: either ≤ 10 prism-dioptre (Δ), i.e. microstrabismus, or with a larger angle of deviation. Finally, children were classified as *strabismic* amblyopes if a manifest interocular misalignment greater than 10Δ was present (convergent or divergent), but no anisometropia.

Exclusion criteria included prior amblyopia treatment other than optical correction, presence of paralytic or restrictive squint, other pre-existing visual deficit (e.g. cataract) or significant neurological or behavioural problems.

3.2.1.2 OT period (eligibility)

One child did not attend the follow up visit #2. For 4/40 children we could not retrieve their baseline-OT acuity. Three of them were then recruited, but 2 did not complete BBV and one, ID1, was released from NHS care much before we analysed the OT data. So, their respective clinical notes were either not accessible or not retrievable at the time of this analysis. Therefore, the following data report OT-results obtained for 35 children.

On average, the OT period lasted 26±10 weeks and the visual acuity in the AE improved from 0.76 logMAR (σ:0.34) to 0.61 logMAR (σ:0.32). Specifically, a total of 9/35 children showed no improvement (change in the AE acuity: μ=-0.17, σ:0.15 logMAR), with 5 of them (14%, of the 35 children) showing a *reduction* greater than 0.1logMAR, considered the acuity test-retest variability (μ=-0.23, σ:0.15 logMAR). In contrast, 20/35 (57%) showed a substantial improvement in the AE acuity (*improvement* greater than 0.1logMAR: μ=0.37, σ:0.19 logMAR); of them, sixteen children showed a mean gain ≥0.2 logMAR (46%; gain in the AE acuity: μ=0.42, σ:0.17 logMAR). The remaining 6/35 children showed only a minor change (≥0.0, ≤0.1logMAR). Overall, two children recovered to within 0.2 logMAR IOAD.

Of the 40 children approached, 24 only could be included in the BBV study (as approved by the ethics). Among the other 16 children, 6 were prescribed patching-therapy at the end of visit #2 as it was difficult to communicate with the responsible optometrist ‘on time’ and so we could not suggest our alternative treatment. Other 2/16 children did not complete their OT period (as mentioned in the first paragraph of this section) and so were not considered further in the study. Of the remaining 8/16 children, there was no record about why six preferred not to participate, while two preferred not to take part into BBV (their parents opted for atropine-drops instead).

3.2.1.3 Recruited participants: details

Twenty-four children (10 males) were finally recruited, after completing their OT period. They aged 3.5 to 11.3 yrs (μ: 6.6, σ:2.9 yrs) and their amblyopia was classified as anisometropic (n=8; Group 1), strabismic or combined mechanism amblyopia (n=6 and 10; both in Group 2). Note that BBV treatment was allowed for <u>maximum 8 weeks</u>

(**Group 1**; pilot study – see Discussion) or maximum 24 weeks (**Group 2**). Participants' details are summarised in Table 4.

The mean period of optical treatment for the 24 children recruited was 28 weeks (σ:12), and specifically 30 weeks (σ:16) for children with anisometropia, 25 (σ:9) for those with strabismus and 29 (σ:12) for those with combined-mechanism amblyopia. The mean gain in acuity during OT period was 0.16logMAR (σ: 0.26), specifically equal to 0.2logMAR (σ: 0.36) for anisometropic children, 0.08logMAR (σ: 0.11) for strabismics and 0.17logMAR (σ: 0.32) for children with combined-mechanism amblyopia. Overall, the mean proportion of deficit corrected was 0.15logMAR with a mean residual IOAD of 0.71logMAR (σ: 0.3), thus reduced from baseline (was on average, 0.76logMAR).

For BBV baseline vision assessment (following visit #2, at ~8 weeks distance), four children were specifically referred to RDCEC from the Moorfields' partnered local services based in Homerton University Hospital (3/4) and in Bedford Hospital (1/4). The other 20 children received their follow-up appointments (details in section 3.2.4) at RDCEC - and so will be for all the 24 children through the course of the therapy.

Our research followed the tenets of the Declaration of Helsinki and we obtained informed consent from caregivers and assent from children prior to enrolment. Our recruitment procedure and treatment regimen were approved by the local NHS Research Ethics Committee.

Participant ID (m/f)	Age (yrs)	Group	Amblyopia		Spectacle prescription	Best corrected visual acuity (logMAR)
			type	severity		
1 (f)	11 .3	1	RE		+0.50/-0.25 x 180	-0.20
			LE aniso (hyperm.)	mod	+5.25/-1.50 x 15	0.36
2 (f)	6.2	1	RE		+4.00/-1.00 x 90	0.10
			LE **aniso** (hyperm.)	sev	+7.00/-0.50 x 10	0.70
3 (m)	10.8	1	RE		+4.50 DS	-0.20
			LE aniso (hyperm.)	mod	+6.50/-0.50 x 30	0.36
4 (f)	10.2	1	RE **aniso** (hyperm.)	sev	+6.50/-3.50 x 15	1.20
			LE		+0.25/-0.25 x 180	0.06
5 (f)	9.4	1	RE		plano/-0.25 x 180	-0.20
			LE aniso (hyperm.)	mod	+4.75/-0.75 x 10	0.42
6 (m)	7.5	1	RE **aniso** (hyperm.)	sev	+2.50/+0.50 x 90	0.68
			LE		plano	-0.04
7 (f)	10.8	1	RE		+0.75/-0.25 x 140	-0.06
			LE aniso (hyperm..)	sev	+7.00/-2/50 x 50	0.64
8 (f)	10.9	1	RE aniso (high myopia)	mild	-8.25/-1.75 x 175	0.28
			LE		-2.50/-2.00 x 3	0.06
9 (f)	4.3	2	RE		+3.75 DS	0.06
			LE comb. (ET- astig.)	sev	+5.5/-0.75 X 175	1.1
10 (m)	5.9	2	RE comb. (ET - hyperm)	sev	+3.00 DS	1.1
			LE		+2/+0.5 X 180	-0.1
11 (m)	5.0	2	RE		+4.75 DS	0.0
			LE strab. (ET)	mod	+5.25 DS	0.38
12 (f)	4.0	2	RE		+3.00 DS	0.1
			LE comb. (ET - hyperm)	sev	+4.00 DS	0.78
13 (f)	7.6	2	RE comb. (ET - hyperm)	sev	+5.5 DS	0.625
			LE		+3.00 DS	0.0
14 (f)	4.4	2	RE		+3.75/-0.5 X 180	0.2
			LE comb. (ET - hyperm)	mod	+4.25/-0.5 X 10	0.56
15 (f)	4.6	2	RE strab (ET ET)	mod	+3.75/-1.5 X 180	0.36
			LE		+3.00/-1.00 X 180	0.02
16 (f)	5.5	2	RE		+6.75 DS	0.06
			LE strab. (ET)	sev	+7.00/-0.50 X 90	1.1
17 (f)	4.0	2	RE comb.(ET – hyperm)	sev	+6.00 DS	1.35
			LE		+1.50 DS	0.15
18 (m)	4.3	2	RE comb. (ET)	sev	+7.50/-1.750 X 180	1.2
			LE		+1.75 DS	-0.18
19 (f)	4.3	2	RE		+4.50/-0.50 X 95	0.14
			LE comb. (ET - hyperm)	sev	+8.75/+1.25 X 90	1.2
20 (m)	4.5	2	RE strab. (ET)	sev	+5.5- DS	1.0
			LE		+5.5 DS	0.1
21 (m)	4.9	2	RE comb. (CShyperm.astig)	sev	+6.50/-3.25 X 10	0.80
			LE		+2.50/-1.00 X 170	0.10
22 (m)	3.7	2	RE strab. (ET)	sev	+2.50/+1.00 X 90	1.1
			LE		+2.50 DS	0.18
23 (m)	11.2	2	RE strab. (ET)	mod	+0.50/-0.75 X 25	0.5
			LE		+0.75/-1.00 X 180	0.0
24 (m)	3.5	2	RE		+4.00/-0.50 X 110	0.125
			LE comb. (ET – hyperm)	sev	+6.00/-0.50 X 160	0.825

Table 4 Baseline details of BBV participants (N= 24). Values corresponding to the amblyopic eye (AE) are italicised. Bold indicates "severe" cases (i.e. acuity in the AE >0.6 logMAR; N=16). Only participant 8 had mild amblyopia, the remaining had moderate amblyopia (0.3≤acuity in the AE≤0.6 logMAR; N=7). Participants 7 and 21 (highlighted) did not attend their clinic appointments, hence their data is incomplete and has not been analysed. m/f=male/female; RE=right eye; LE=left eye; aniso.=anisometropic; strab.=strabismic; comb.=combined mechanism; hyperm.=hypermetropia; astigm.=astigmatism; DS=dioptre sphere; ET=esotropia; strabismus; sev.=severe; mod.=moderate

3.2.2 Equipment

Our treatment uses a computer system capable of presenting 3D movies, which is installed in the child's home (Figure 18A). The monitor operates at 1920x1080 pixel resolution at 120Hz (60Hz per eye). Movies were presented using software written in MATLAB (MathWorks, Ltd., Cambridge, MA) and Psychtoolbox (Brainard, 1997). Shutter glasses (nVidia Corp., Santa Clara, CA) were used to independently control the image presented to the two eyes. These were mounted in a customised children's ski mask to ensure comfort, while maintaining a snug fit over their spectacle correction. The monitor was linearized in software based on a series of luminance measurements (made by placing a Minolta photometer behind a single lens of a pair of goggles) to achieve a mid-grey of 45cd/m^2 to each eye. Children were provided with a keypad to make responses to the suppression task, and were encouraged to use the 95cm long cable to ensure they maintained viewing distance at around 1m.

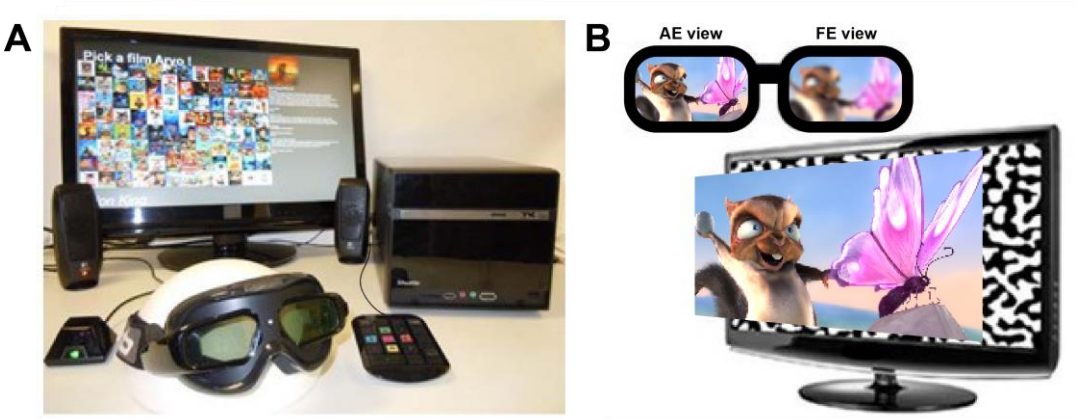


Figure 18 A. BBV treatment system: a PC, 3D-capable monitor, response-keypad, IR emitter and goggle-mounted shutter-glasses. **B.** (top) The child's view through amblyopic (AE) and fellow (FE) eyes. The system applied sufficient blur to the FE to match acuity with the AE. (bottom) Although movies are 2D, a shift in the relative position of each eye's image modulates the perceived depth of the movie. The zero-disparity textured background is also visible, which provides a vergence lock.

3.2.3 Treatment regimen

Treatment consisted of 1 hr/day spent viewing movies (selected by children/carers) while wearing the goggles. Movies were presented dichoptically and the horizontal offset between the two eyes was continuously modulated to generate a percept of gradually changing depth. A zero-disparity textured background was presented to both eyes to encourage stable vergence. The child's view of the movie (Figure 18B, inset) was "balanced" by blurring the image presented to the FE, so that the child's monocular acuity was matched across eyes. To determine the level of blur required, we ran two tasks. Task 1 (Figure 19A) quantified AE acuity as the scaling required to support identification of the (mouth) orientation of a crowded Visual-Acuity-Man ("VacMan"; Greenwood et al., 2012). Targets were presented monocularly at 75% contrast with four flanking "ghosts" (spaced at twice the target-width). They were masked with a 25% contrast phase-scrambled version of the stimulus that was visible across both eyes (to provide a vergence lock). Stimuli were scaled using an adaptive staircase (QUEST; A. B. Watson & D. G. Pelli, 1983). Over 45 trials this converged on the scaling that produced 83% correct identification. Task 2 (Figure 19B) presented similar VacMan stimuli to the FE (scaled with the AE threshold from Task 1) and then used QUEST to determine the level of isotropic Gaussian blur that elicited 83% VacMan-identification. This level of blur was applied to the image presented to the FE during movie presentation, ensuring that the images presented to the two eyes were equally visible.

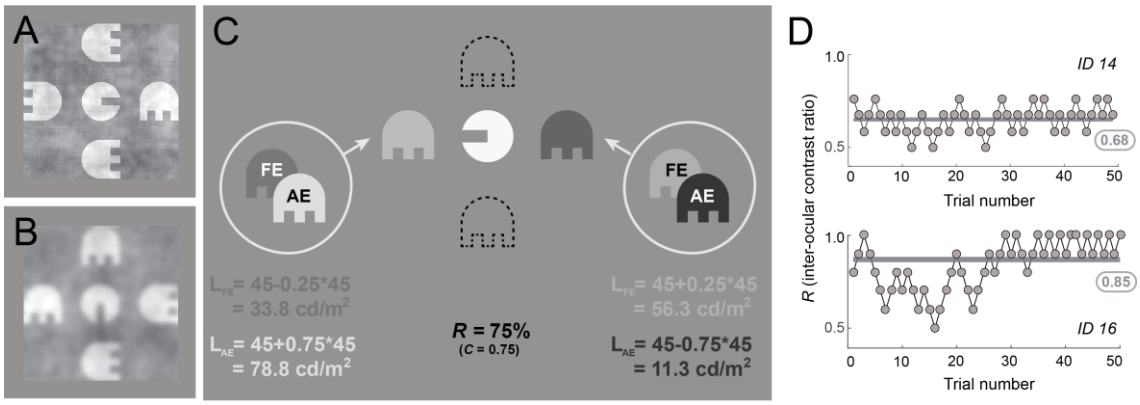


Figure 19 Setting the binocularly balanced vision.. The psychophysical tasks used (A, B) to establish blur level for the FE and (C) to quantify suppression during treatment. **A.** Crowded "VacMan" stimulus used to measure acuity in the AE. **B.** Stimuli used to estimate the blur level required to match the performance of the FE to the AE. **C.** Suppression/compliance task. VacMan is flanked by two ghosts either positioned on the left and the right (as shown) or above and below (dashed outlines). Each ghost was a mixture of one dark and one light ghost presented to different eyes on each side (illustrated within the white circles). We quantify suppression as the mixture of luminance (L) -increments and -decrements required for the child to be equally likely to report either ghost as "whiter". **D.** Sample staircases for two observers. Grey text and horizontal line indicate the estimated balance point.

During treatment, the movie was interrupted every minute by an interactive game used to measure suppression (Figure 19C). Two dichoptically presented "ghosts" flanked a central VacMan, either above/below or left/right. We told children *"VacMan wants to eat the whitest ghost – which ghost looks whiter?"* They responded (up/down/left/right) using a keypad. Each ghost was composed of one dark and one light component, presented dichoptically to each eye. The luminance of the components was set using an inter-ocular contrast-ratio (*R*; 0-100%), which determines the relative strength of FE and AE stimulation as 0%= FE suppressed, 50%=balanced vision and 100%=AE suppressed**.** For a background luminance of L_{back} (here, 45cd/m^2) with a maximum increment/decrement of L_{range} (here, 45cd/m^2) we made stimuli with this algorithm:

1. Randomly select if ghosts fall above/below or left/right (Figure 19C) of VacMan
2. Randomly assign the light/dark FE/AE polarity ghost to one side of VacMan, with the opposite polarity on the other side, e.g. Fig. 2C shows a dark/light FE/AE ghost on the left and a light/dark FE/AE ghost on the right of VacMan.
3. Use *R* to set ghost component-luminance values. The luminance of the light ghost on one side of VacMan and the dark ghost on the other side are matched increments and decrements: $L_{back} \pm (R/100)*L_{range.}$ For example in Figure 2C, *R*=75%. Thus, for *C*=*R*/100=0.75: 45± *C* *45 = 78.8 and 11.3 cd/m^2, respectively. Similarly, the light ghost (right) and dark ghost (left) are $L_{back} \pm (1-C)*L_{range}$ =45±0.25*45= 56.3 or 33.8 cd/m^2.

Initially *R* was set to 80% and then adjusted (by ±10%) on each trial, using a one-up-one-down staircase procedure, according to whether a response was consistent with reliance on the FE or the AE. For *R*=75% (Figure 19C) the AE sees a substantially stronger white ghost; someone with balanced vision would report that they perceive the whiter ghost on the "left" leading to a reduction in *R* on the next trial. Thus, *R* converges on the level necessary for the stimulus delivered to either eye to drive the report with equal probability. For an observer neglecting the AE, *R*>50% indicating that a stronger signal is required in the amblyopic eye in order for that ghost to be reported as whiter.

Note that each child performed one trial/minute with the procedure restarting each time the child switched movies, so the number of trials contributing to any one suppression-estimate varied according to the time spent watching a given movie in one session. If the child chose a location where no ghost was presented, he/she was either not paying attention or not viewing through the goggles (since balanced ghosts were invisible without the 3D shutter glasses). The number of such errors was used to quantify attention/compliance. The system emailed the experimenter a daily update of the time children had spent engaged in movie viewing and their performance on this task.

We treated the first cohort of participants (Group 1: N= 8, all anisometropes) for a maximum of 8 weeks. At standard orthoptic assessments that occurred alongside BBV (see next section), we observed gains in acuity and stereoacuity that did not reach a plateau. Therefore, for the second group of children (Group 2: all combined or strabismic amblyopia) we extended the maximum period of treatment to 24 weeks.

3.2.4 Orthoptic Assessment

We based the evaluations of the treatment's safety and efficacy on the results (primarily, clinical acuity) obtained during a standard orthoptic assessment. Based on these results, BBV treatment could have a different duration also between children included in the same group, either Group 1 or Group 2 (respectively, allowed for the same maximum duration).

One of the collaborating experienced orthoptists performed a series of clinical tests at baseline (pre-treatment) and after 4 and 8 weeks of therapy, for every child, and at 12, 16 and 24 weeks for others (only those included in Group 2), depending on the individual response to treatment. The orthoptists assessed:

- best-corrected visual acuity using a crowded logMAR test at 3m distance (Thompson v2000 software; Thompson Software solutions, Hertfordshire, UK) – that we considered as our primary outcome measure;
- stereoacuity, using the Frisby near stereotest (Anketell, Saunders, & Little, 2013)
- ocular motility and ocular alignment at 3m (distance) and 33cm (near) using the prism cover test.

They also attempted to make a clinical measure of suppression, as an additional safety evaluation of BBV, using Bagolini filter bar (aka Sbisa bar). In fact, this test uses red filters of increasing density to quantify the reduction in luminance of a target (presented to the fixating eye) required to induce diplopia. However, this test was difficult to administer (only seven of the children approached were able to perform it). Given the small sample-size, and reports of poor test-retest reliability of this test (M. Piano & Newsham, 2015), we do not consider these data further.

Participants eligible for a longer course of treatment (Group 2; see *Results*) were also assessed at 16 and 24 weeks. Treatment was discontinued after 4 weeks if the acuity fell below baseline, or if the inter-ocular acuity difference (IOAD) had improved to normal levels (0.2 logMAR or less). At subsequent visits, children were considered to have reached a plateau if acuity failed to improve by at least 0.1 logMAR from their preceding

visit. Children who were advised to discontinue home-therapy were referred back to the hospital eye clinic to receive standard occlusion therapy, if IOAD was still 0.2 logMAR or greater or AE acuity did not recover to within 0.1 logMAR.

3.2.5 Outcome measures

We used crowded logMAR acuity to evaluate the efficacy of BBV treatment. We expressed changes in visual function as:

- AE logMAR acuity (our primary outcome measure);
- residual IOAD after treatment;
- proportion of deficit corrected, defined as $(AE_{baseline}-AE_{exit})/(AE_{baseline}-FE_{exit})$ (Stewart, Fielder et al. 2005).

In addition, we explored stereopsis (Frisby test) and suppression (ghost task, described above). We also quantified compliance as the mean time spent watching movies per day ("daily dose") and the mean cumulative time spent watching movies ("total dose") and adherence as the percentage of days treatment was received. Other factors that may have contributed to the outcome measures (treatment duration, type of amblyopia, initial severity of amblyopia and age) were also evaluated (Stewart et al., 2005) and are described in section 3.3.4.

3.3 Results

Two children did not attend the 4-week appointment and were excluded from analyses leaving 22 children. Group 1 thus consisted of 7 children with anisometropic amblyopia (mean age 9.5 yrs; 4females). A total of 15 children were included in Group 2 (mean age 5.2 yrs), 6 with strabismic amblyopia (mean age 5.75 yrs; 2females) and 9 with combined-mechanism amblyopia (mean age 4.7 yrs; 6females). Where necessary, results are reported separately for children included in Group 1 (allowed max 8 weeks) or 2 (allowed up to 24 weeks on treatment) – see also Table 5.

3.3.1 Acuity

Figure 20A and Figure 20B plot the difference in logMAR acuity from baseline (BL), as measured for each child during his/her clinical assessments. Specifically, data for children with pure anisometropic amblyopia (N=7; square symbols) or strabismic amblyopia (N=6; triangle symbols) are reported in Figure 20A, while those with combined mechanism amblyopia are in Figure 20B (N=9; circle symbols). Individual declines or improvements in vision are represented by values falling above or below the

dashed horizontal line respectively (no change from BL = 0 logMAR acuity difference). The individual values measured before starting and after completing BBV treatment (entry vs exit logMAR acuity) are plotted in Figure 20C. As per the study protocol, children were treated for up to either 8 weeks (Group 1; N=7: IDs 1-8 in Table 1) or 24 weeks (Group 2; N=15: IDs 9-24). We summarise in

Table 5 the different duration of treatment resulting for each child, depending on study protocol and the orthoptist's decision (also based on the protocol) at each programmed visit. Among children in Group 2, five did not improve further after 8 weeks (IDs: 9, 10, 17, 19 & 24; making 12 children in total released at this time-point), while 4 children remained in treatment for 16 weeks (IDs: 11, 16, 18 & 23) and 6 for 24 weeks (IDs: 12-15, 20, 22), depending on the measured improvement in acuity. Note that acuity continued to improve beyond 8 weeks for some children, suggesting that those whose treatment was terminated at this point (because this was the limit of our approved protocol for Group 1) could have received further benefit from continued treatment.

	Participants' IDs																							
Did not attend 4w. visit							7													21			
Stopped @**8** weeks	1	2	3	4	5	6		8	9	10							17		19					24
Stopped @**16** weeks											11					16		18					23	
Stopped @**24** weeks												12	13	14	15					20		22		
	Group 1								**Group 2**															

Table 5 Allowed and actual individual BBV duration. Treatment duration is indicated for children included in Group 1 (IDs 1-8; in red, to indicate the common type of amblyopia: anisometropia) and in Group 2 (IDs 9-24; in yellow or blue, respectively indicating children with strabismic or combined-mechanism amblyopia). As indicated by the two grey arrows, treatment duration was fixed at maximum 8 weeks for Group 1 or maximum 24 weeks for Group 2. None of the children was released during the first follow up, at 4 weeks after recruitment, to respect our safety/efficacy measures, but two children did not attend and so were released.

When data from all the children are combined (regardless of amblyopia type, N=22), mean acuity in the AE improved from 0.78 ±0.35 to 0.51 ±0.34 logMAR, a significant mean gain of 0.27 ±0.22 logMAR (one-sample paired *t*-test, $t_{5.83}$, $p<0.001$). Vision in the FE remained stable, improving slightly: the mean gain (-0.05±0.11 logMAR, mean baseline: 0.02 ±0.13logMAR) was statistically ($p=0.04$) but not clinically (gain<0.2logMAR) significant. Mean acuity gain in severe amblyopia (indicated as 'sev.' in Table 4 for the N=14 with acuity in the AE worse than 0.6 logMAR) was 0.32 ±0.24 logMAR, vs. 0.18 ±0.14 logMAR in mild-to-moderate amblyopia (N=8).

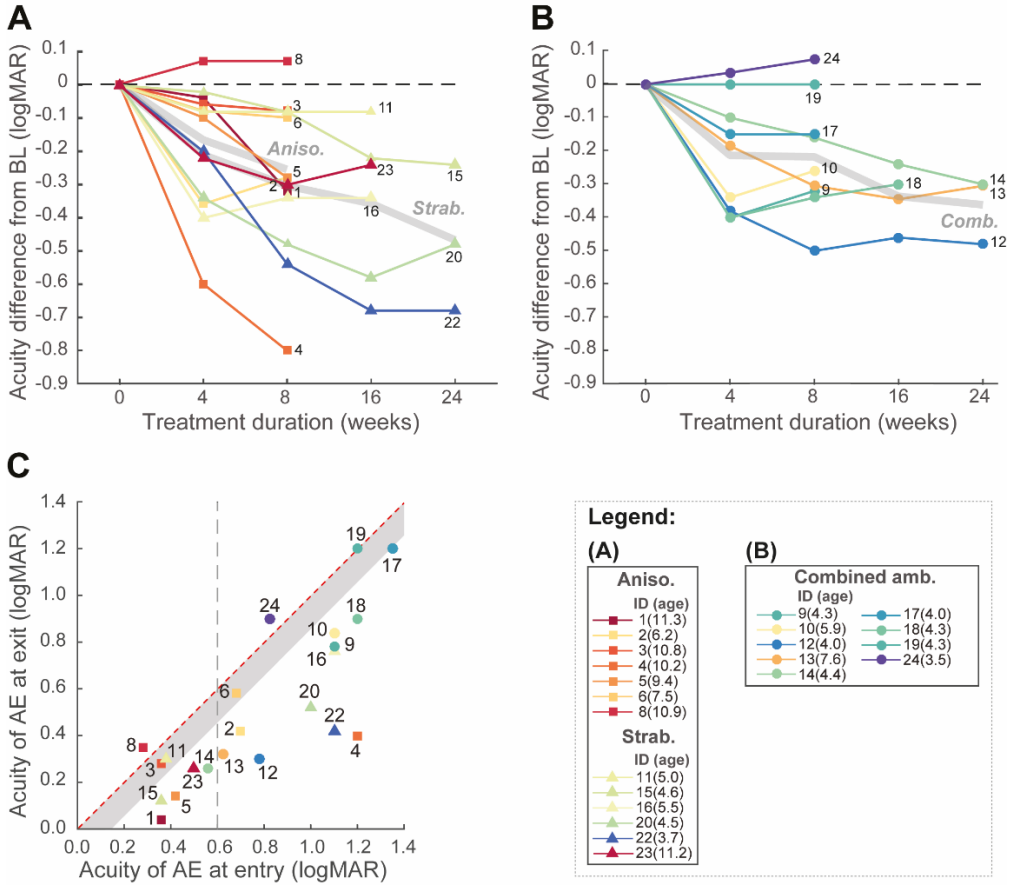


Figure 20 A, B. Acuity difference in the AE compared to baseline (BL) during treatment; changes below or above the dashed line respectively represent improvements or deteriorations in vision. Participants' amblyopia-type was pure anisometropic (N=7; A: squares), pure strabismic (N=6; A: triangles) or combined (N=9; B: circles). Thick lines show the mean change in acuity-difference for each type of amblyopia (aniso. and strab. in panel A; combined amb. in B). Symbol-colour codes the age of participants (age indicated in parentheses, in legends of parts A,B; from blue-younger to red-older children). Identity-codes (1-24) are given next to individual lines and in the legends of A and B and label individual data point in C. Note that Group 1 (A) are anisometropic (i.e. treated for a maximum of 8 weeks), and Group 2 (A, B) are strabismic and combined (treated for a maximum of 24 weeks; line length gives length of treatment). **C.** Comparison of pre- and post-treatment acuity for each child. Points below the diagonal line are improvement, with the shaded region indicating gains less than 0.15 logMAR (considered critical of test-retest reliability (Tsirlin et al., 2015). The dashed vertical line (at 0.6 logMAR) represents the cut-off between mild-to-moderate and severe amblyopia.

Figure 21 replicates 20C but using different symbols, to easily visualise the improvement based on the different types of amblyopia (see section 3.3.4 for more details) and to compare the acuity gains for Group 1 and 2. Respectively, for Group 1 (whose treatment was curtailed at 8 weeks, all anisometropic amblyopes) the mean acuity gain was 0.26 ±0.28 logMAR, and for Group 2 (maximum treatment of 24 weeks, combined and strabismic amblyopes) it was 0.27 ±0.19 logMAR. There was no significant difference in acuity gains between Group 1 and 2 (two-sample t-test$_{(df:20)}$, p=0.863).

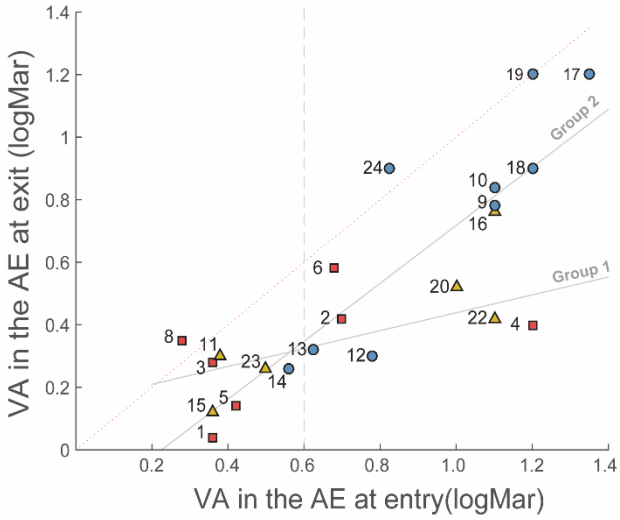


Figure 21 Acuity in the AE at exit compared to baseline (at entry) for children who completed BBV (the red dotted line is for identical entry-to-exit acuity; any symbols falling below this line indicates an improvement). Different symbols identify the different types of diagnosed amblyopia: red squares for anisometropic, yellow triangles for strabismic and blue dots for combined-mechanism amblyopia. The vertical dashed line marks the lower limit for severe amblyopia cases (i.e. VA in the AE at entry > 0.6 logMAR). Two least-squares lines (in grey) highlight the correlation between x and y-values, for Group 1 (R=0.508, *p*=0.245) and Group 2 (R=0.846, *p*=0.0001).

After treatment 15 children reached IOAD ≤ 0.6 logMAR (6 of whom started with severe amblyopia), including 7 children (one severe) who recovered to ≤ 0.3 logMAR. No further treatment was required for ID15, whose IOAD improved from 0.34 to 0.1 logMAR and ID1, whose AE acuity reached 0.04 logMAR (although final IOAD was 0.24 logMAR). The mean "proportion of deficit corrected" was 32 ± 26%, with substantial gains (>60%) in 2 children (IDs 15 and 22, 71% and 69%), and poor (<10%) in 3 children (IDs 8, 19 and 24). In seven children, improvement was between 10 and 30%, and in the remaining 10 children, between 30% and 60%.

3.3.1.1 Stereoacuity

Only children with purely anisometropic amblyopia (Group 1) had measurable stereoacuity at baseline, with a median of 170 arc sec (interquartile 230 arc sec; Figure 22A). Following treatment, their median stereoacuity value was 85 arc sec (interquartile 30 arc sec; N=7). Among the 7/8 children completing BBV, one (ID3, who had good stereoacuity at entry) showed no stereoacuity improvement, while the other 6 showed a mean gain of 165±182 arc sec. This demonstrates the possibility for BBV to be effective on stereoacuity. We transformed data to logarithmic seconds of arc to calculate if "real change" in stereoacuity occurred. Prior studies have found the test-retest reliability of stereoacuity measurements using the near Frisby test in children to be 0.3 log arc sec, with "real change" defined as a doubling of stereoacuity expressed in octaves (Adams, Leske et al., 2009). Here, mean stereoacuity gain was 0.40 log arc sec (σ:0.32), with all

but ID3 exhibiting an improvement in stereoacuity ≥1 octave (Figure 22B). Mean improvement was 1.33 octaves (i.e. a factor of 2.6 improvement). The gain in stereoacuity significantly correlated with both the initial level of acuity in the AE (r=0.97, p=0.0003) and its absolute improvement (r=0.85, p=0.02), but did not correlate with the proportional gain in acuity (r=0.44, p=0.32).

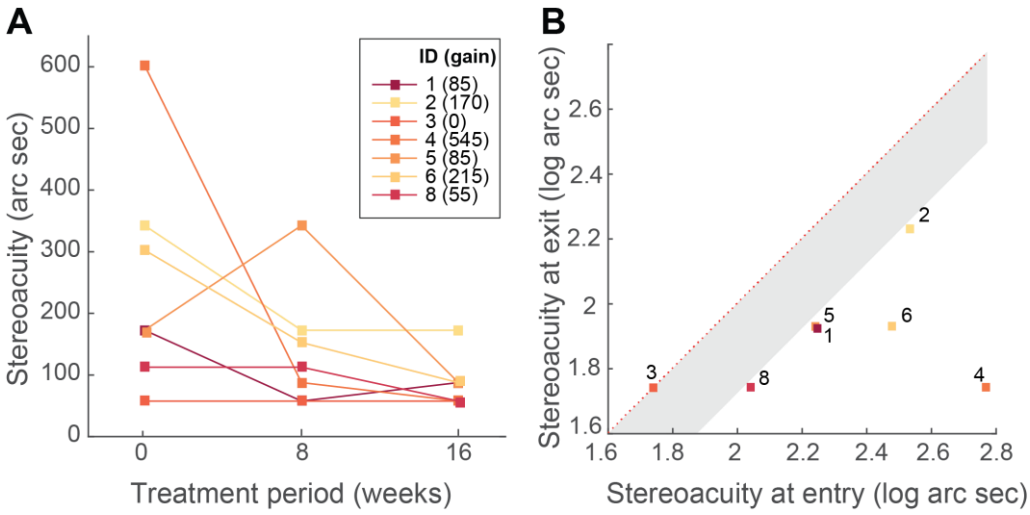


Figure 22 Stereoacuity (when measurable) for children who completed BBV treatment. **A.** Before, during and after treatment (at 0, 8 and 16 weeks respectively), **B.** Pre- versus post-treatment. All children for whom data are shown (N=7) had purely anisometropic amblyopia (Group 1). Symbols are coloured to reflect the relative age of each child compared to their peers (red=oldest). Boxed legend in A shows individual gain in stereoacuity (arc sec). For 6 participants stereoacuity gains exceeded the test-retest variability threshold of 0.3 log arc sec and a step of one octave (shaded area).

Of the children having unmeasurable baseline stereopsis, one showed progression at the end of BBV treatment (ID14: 600 arc sec) and three at the ~1 year follow up, reaching 600 (ID9), 85 (ID11), 300 (ID12) – when ID14 further improved to 110 arc sec. Overall, the results on half of the children gave us the possibility to consider BBV as potentially effective on stereoacuity (depending on the initial severity and type of amblyopia) but the clinical notes for the other 50% reported 'inconclusive' thus we were not able to discuss Frisby data further.

3.3.2 Interocular suppression

Figure 23 shows individual suppression data from the ghost task. Note that ID4 initially did not comply with this task and the relative (incomplete) data were excluded from the analyses of suppression. The mean suppression at entry was 72.3% (σ: 12.02%) and at exit was 72.6% (σ: 12.3%). Overall, these values are (a) in line with comparable estimates for adult amblyopes (e.g. 75%; Kwon et al., 2015) and (b) not significantly

different from one another (t-test$_{(df:20)}$: p=0.98). We do not observe the substantial reductions in suppression observed in other studies of binocular therapy (Vedamurthy, Nahum, Bavelier et al., 2015). Indeed, a statistically significant reduction in interocular suppression was observed in only 6 of the 22 children, of whom 4 had combined-mechanism and 2 purely strabismic amblyopia. Further, 5 children showed a significant *increase* in suppression (ID4 excluded) while 10 children showed no significant change. For each individual we calculated the linear regression trend-line (bold lines in Figure 23) for daily estimates of R (quantifying the patient's binocularity: 0% fully reliant on AE, 100% fully reliant on the FE).

We performed additional analyses to examine changes within and across sessions. First, for within-session changes, we analysed runs containing at least 30 trials ('long sessions'; an average of 36.5% of all runs across 21 children) and divided these runs into three parts. We then compared the average stimulus balance in the second and third part (excluding the first part where the staircase may not be close to convergence), computing a linear regression between values to determine if the slope was significantly different to 0. According to this analysis, for each child (excluding ID4) an average of 9% of 'long sessions' involved a significant change in suppression within session. However, such changes were not biased towards increasing suppression (49.8±27.8% of cases) or decreasing suppression (50.3±27.8%). Second, across sessions, we note that Kehrein, Kohnen, and Fronius (2016) have reported an increase in suppression during the first thirty days of occlusion therapy, followed by a return to baseline in the following month. We performed a similar analysis comparing suppression over the first and second 30 days of treatment using regression analysis. Mean slopes over the first and second ~30 days were 0.0688 (σ: 0.4220) and 0.0018 (σ: 0.6245) respectively, a non-significant difference ($t_{(20)=0.41}$, p=0.68). Thus, occlusion may exert greater (but short-lived) influence on inter-ocular suppression than binocular therapies.

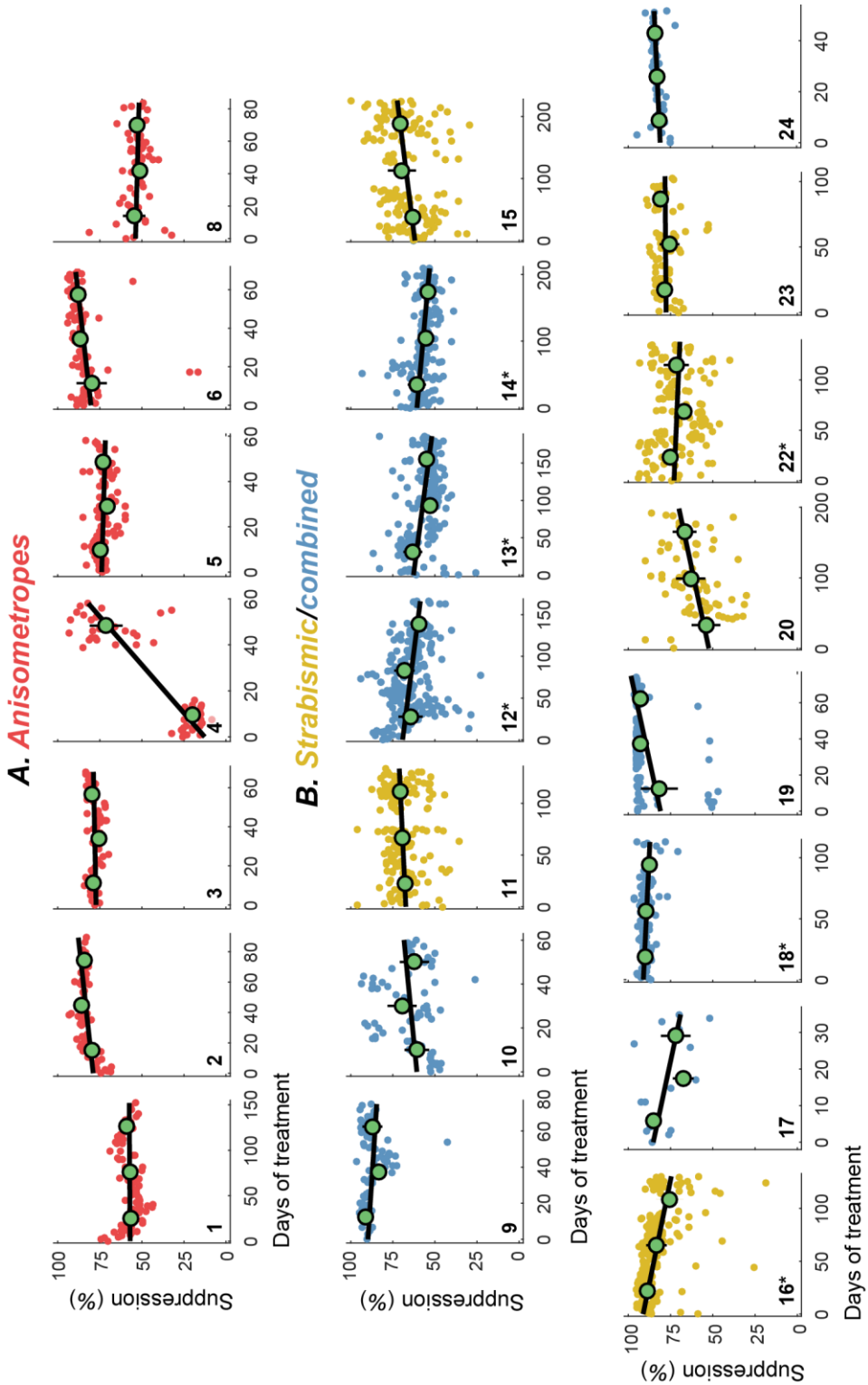


Figure 23 Daily estimates of suppression (*R:* % reliance on fellow eye, *see Treatment regimen*) for participants with anisometropia (A.: Group 1) and/or with strabismus (B.: Group 2). Participants' IDs numbers are indicated at the bottom left of each subplot. Here, 50% means "balanced vision", 100% indicates complete reliance on the FE (i.e. complete suppression of the AE), and 0% indicates complete suppression of the FE. Green symbols pool data within three periods (beginning, middle, end; binned around the individual duration of BBV; note ID4 was not compliant for a period, hence the middle bin is missing). We derived black trend lines from linear regression analysis of daily estimates. An asterisk after the child's ID indicates a significant reduction in suppression for that child ($p < .05$).

Figure 24A shows suppression at entry versus exit from treatment (ID4 did not have a complete set of data and was excluded). There was no systematic trend in the change of suppression with treatment – suppression decreased in some children (points below the unity line) but increased in others (points above unity). Figure 24B plots improvement in acuity for the AE versus the difference in suppression, obtained by averaging each child's daily suppression measures. There was a *non*-significant tendency for more improvement in acuity to be associated with modified suppression (Pearson's r=0.19, *p*=0.40; ID4 excluded), especially when suppression significantly changed - either increasing (r=-0.47; *p* =0.80) or decreasing (r=-0.13; *p*=0.35). For observers with stable suppression (N= 10) we observed a mean gain in acuity of 0.16 ±0.15 logMAR, whereas for observers whose suppression decreased (N= 6) the change in acuity was 0.40 ±0.15 logMAR and for those whose suppression increased (N=5; ID4 excluded) the change was 0.22 ±0.18 logMAR. In Figure 24B we highlight the range of uncertainty for each child by adding error bars (denoting 95% confidence intervals; horizontal bars for acuity-gain, verticals for change in suppression). To do this, we first estimated confidence on acuity gain from the typical test-retest variability of logMAR acuity results in children with amblyopia (±0.15logMAR; the typical test-retest variability for logMAR charts) and on change in suppression by resampling our indices. We then bootstrapped on pairs of derived values, obtaining at each repetition two sets of changes to find the relative correlation (one set from acuity paired-values and one from suppression paired-values). The mean *r* across repetitions was 0.161 (σ: 0.132) confirming the lack of a strong correlation.

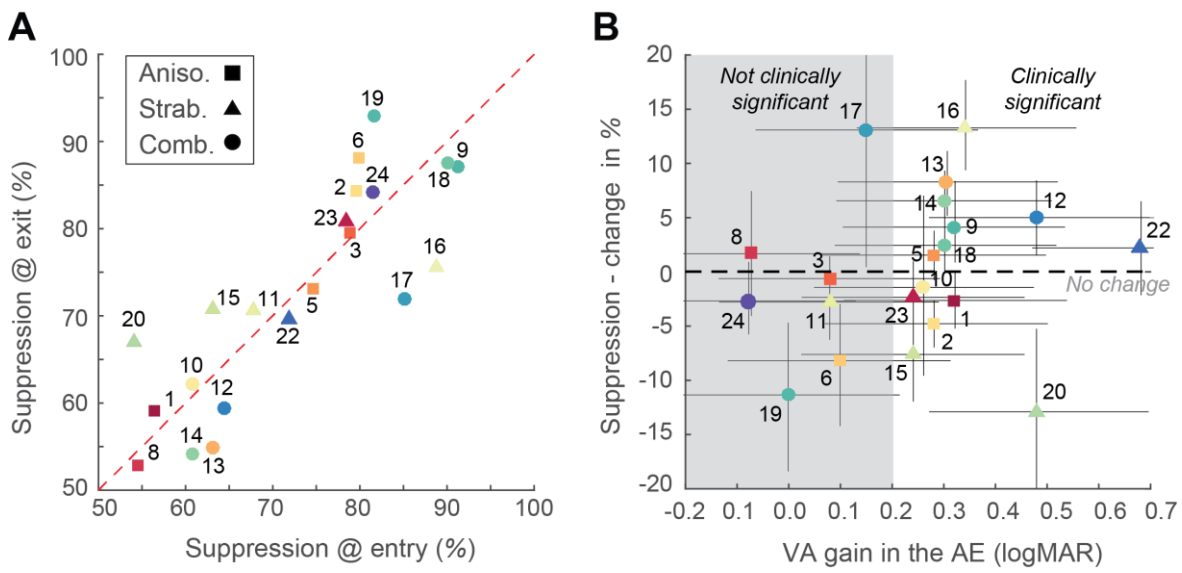


Figure 24 Change in suppression and the relation to baseline acuity.Estimates of suppression for 21 children (ID4 excluded; IDs numbers to label the correspondent data points; markers coloured from blue-younger to red-older children across all recruited children) **A.** Suppression (*R*) is similar at entry compared to exit from treatment. **B.** There is only a modest correlation of gain in visual acuity (VA) with change in suppression (r=.193) which does not reach statistical significance (*p*=.402). The horizontal dashed line represents the level of "balanced vision". VA gains outside of the shaded region are considered clinically significant (>= 0.2logMAR). Error bars indicate 95% confidence intervals on acuity (horizontal) and suppression (vertical) estimates.

3.3.3 Compliance

On average, adherence (calculated as the percentage of days where treatment was received) was 68±12.2%, meaning that children watched a movie on more than two-thirds of the days on which the equipment was available to them. The mean total dose (across the whole treatment duration) was 75 hours 14 minutes, with a mean daily dose of 54 ±14.5 minutes (range 25 to 89 min). Figure 25A shows that none of the children used the system for less than 20 minutes, on average, per day (30% of the prescribed dose). Good compliance (20-50 min) was demonstrated by 7 children, and excellent compliance (>50 minutes a day) by 15 children, with mean adherence of 63.4% and 70.5%, respectively. Five children exceeded the prescribed dose. A previous study on a monocular video-game therapy for amblyopia showed a marginally significant correlation between gain in acuity and longer daily sessions in children (Zahra Hussain et al., 2014). We find that a greater final gain in acuity was not significantly associated with greater daily dose (r = 0.234, p=0.296; Figure 25A) or with a higher percentage of days on treatment (r = 0.0001, p=0.9998). One might expect dedication to therapy to improve with age, but we did not find significant correlations of age either with the daily dose (r=-0.05, p=0.8) or with the number of treatment days (r=-0.38, p=0.08).

As a measure of attention paid to the task, we classified responses on the ghost task either as "valid" (the child indicated a ghost in a location where there was one) or as

“lapses” (the child indicated a position were no ghost was present). On average 23.1% of responses over all runs were “lapses” (σ: 20.7%). Figure 25B shows a significant negative correlation between the proportion of “lapses” and the age of the child (r= -0.54, p=0.01). Although we note a high number of lapses, particularly in some younger children, this does not seem to have greatly impacted our estimates of suppression (which remain stable across many days; Figure 23). In particular, while some children with noisier suppression data did make more lapses (e.g. IDs 15, 22), others did not (e.g. IDs 10, 11). This is because our suppression estimate was tolerant of lapses since (a) lapsing generated “no-response” (not a random response) so that the same stimulus-level was presented until a valid response was made and (b) run lengths were long enough (μ: 30 trials, σ: 14.8 trials) that staircases converged even with frequent lapsing. We note that children generally tolerated the interruption of the “ghost” task surprisingly well, possibly because they are accustomed to media content being regularly interrupted by commercials.

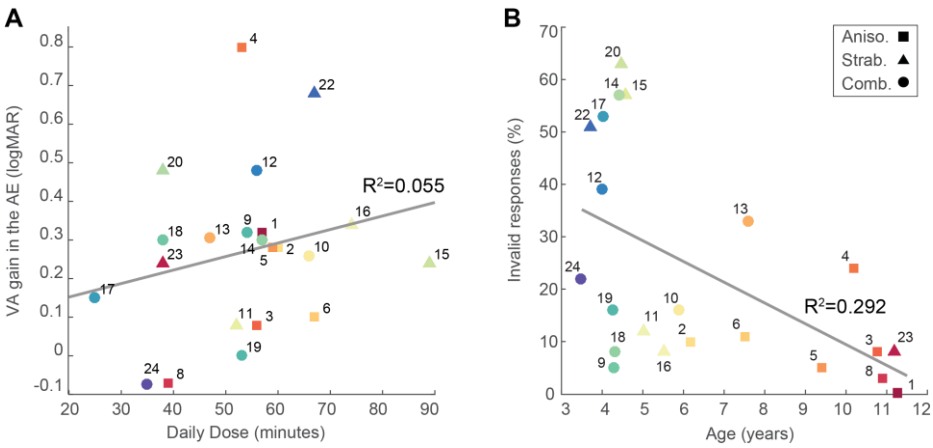


Figure 25 Compliance and attention. **A.** Daily dose in minutes is plotted against acuity (VA) gains in logMAR for the AE. The trend line shows how a higher daily dose (individual mean number of minutes per day spent watching movies) seemed associated with greater improvement in VA. **B.** Percentage of lapse trials on the whitest-ghost task (“invalid responses”, i.e. where the child either indicated a location where a ghost was not present or did not respond at all) as a function of age. Younger children are more prone to lapsing, possibly indicating poorer attention to the task.

3.3.4 Visual Outcome: Other Factors

Figure 26 summarises how treatment duration, type and severity of initial amblyopia and age influenced the outcome (see Table 5 for more individual-details).

Treatment duration. Of the children who stopped after 8 weeks (N=12), 7 from Group 1 were not allowed to continue based on our protocol, while 5 from Group 2 were released due to a lack of further improvement. The mean gain in acuity after 8, 16 or 24 weeks was 0.20±0.24, 0.24±0.11 or 0.41±0.16 logMAR respectively (Figure 26A). Paired

comparisons did not reveal a significant influence of the duration of the treatment on the final gains in acuity (not even between 24 versus 8 weeks, whose relative mean acuity gains showed the largest reciprocal difference; p=0.07). Differences in the maximum permissible period of treatment across groups precludes detailed comparison between groups. However, dependence of treatment response on the type of amblyopia is summarised in Figure 26B.

Type and severity of amblyopia. Visual acuity improved on average by 0.26 (σ: 0.28), 0.34 (σ: 0.21) or 0.23 (σ: 0.17) logMAR respectively in children with anisometropic, strabismic or combined-mechanism amblyopia. A two-sample paired t-test indicated there was no statistical significance between the mean-gain achieved for each type of amblyopia (anisometropic vs strabismic, p=0.54; anisometropic vs combined, p=0.80; strabismic and combined, p=0.26). There was no significant dependence of the severity of initial amblyopia with either the final absolute improvement in vision (mean acuity gain in the AE; R^2=0.13, p = 0.09; Figure 26C) or the final proportion of deficit corrected (R^2=0.0006, p = 0.91).

Age. Lower age has been associated with higher probability of a successful treatment, possibly preventing the applicability of a treatment in adults: compliance (e.g. to patching) reduces with increasing age (Wallace et al. 2013, Stewart et al., 2005, Scheiman et al., 2005) and so does cortical plasticity (Lewis et al., 2005). Using regression analysis (least-squares fitting) we found that the age of participants did not differentially influence the change in acuity in the AE (R^2= 0.001, $p_{(F=0.03)}$ =0.87; Figure 26D). Accordingly, there was no dependence of age for children in Group 1 (mean age 9.46±1.93 yrs) or Group 2 (mean age 5.12±1.97 yrs), on the final gain in acuity in the AE (two-sample t-test: p=0.86). Within Group 1, we did however find an effect of age on stereoacuity measurements (Wilcoxon-paired, p=0.02 at α=0.05), though this was not measurable in Group 2. Finally, we considered the possibility that suppression changed with age by taking the individual suppression index averaged over the daily measures, for the duration of BBV treatment. Here we found no significant dependence of age on suppression (R^2= 0.14, $p_{(F=3.21)}$ =0.09).

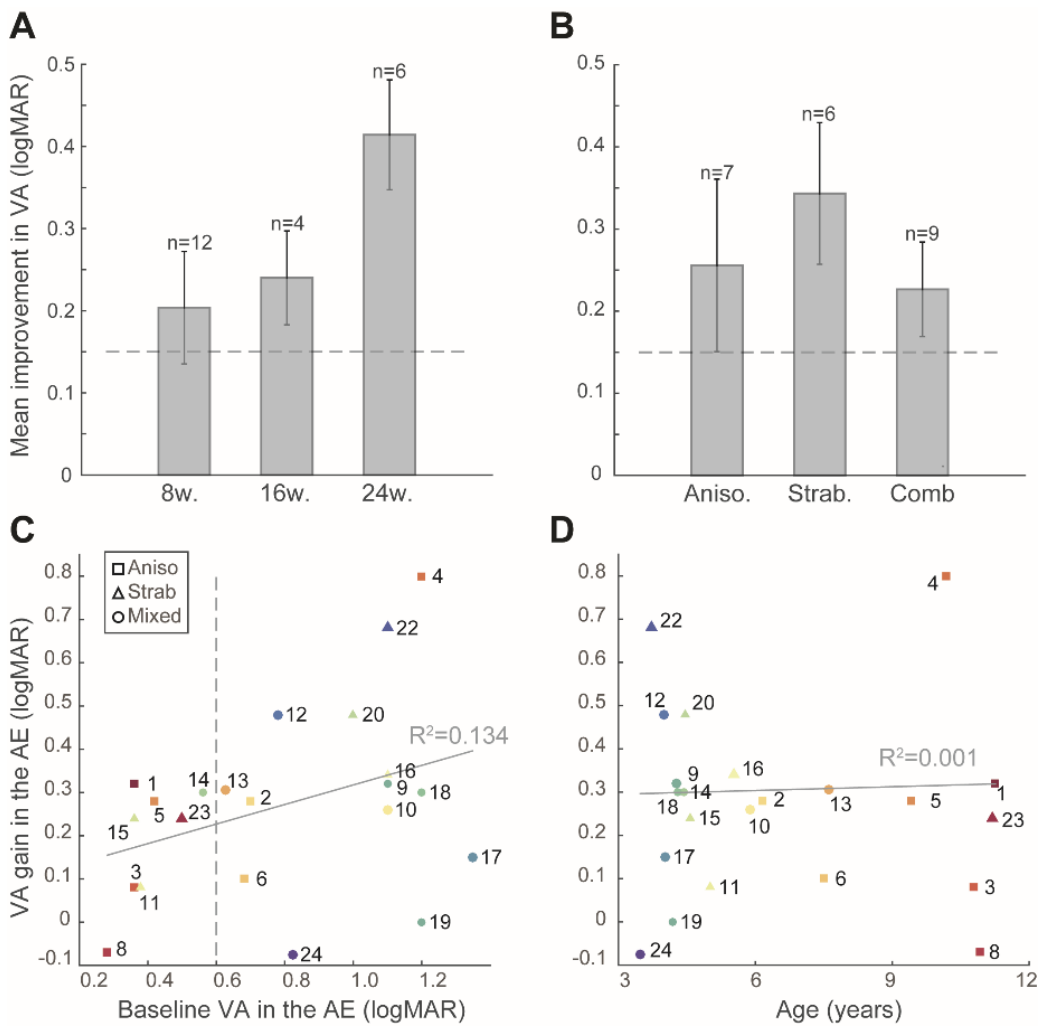


Figure 26 Other factors that may have influenced treatment outcome (A,B: mean improvement or C, D: individual gain). **A**: Longer treatments led to greater but not significantly different gains in the AE visual acuity. **B**: Relations between the type of amblyopia and the improvement in VA (none were significant). The dashed lines (in A and B) show the mean test-retest reliability for acuity tests in children. Error bars show one standard error of uncertainty. **C**: Severity of amblyopia (initial VA in the AE) showed moderate influence on acuity gain in the AE, though this was not significant (R^2= 0.134, p=0.09). **D**: Age (yrs) was not significantly correlated (R^2= 0.001, p=0.87) with acuity gain in the AE.

3.3.5 Maintenance of acuity gains after end of treatment

In February 2018, we re-checked the clinical records for the 22 children who completed BBV. Five children had been dismissed from NHS care (2 did not attend their appointments twice and were discharged, 1 was not compliant to testing twice and was referred to a local optician and 2 were not contactable anymore) and one was not referred yet for a follow-up (possibly, the child moved to private care and/or records were incomplete). Follow-up data were instead available for the remaining (16/22) children. Of them, 13 received patching treatment (on average for 3hrs/day) following BBV. Specifically, 2/13 were prescribed additional atropine-drops twice a week, to try compensating their very low compliance to the prescribed 6hrs daily dose of patching. For the 16/22 children, the mean gain in the AE acuity was 0.32 ±0.21 logMAR at

completion of BBV treatment (in particular, the acuity gain was =0.32±0.22 logMAR for those who then received occlusion therapy). The mean acuity measured on their AE improved from 0.51±0.31 logMAR, at treatment completion, to 0.39±0.33 logMAR, after about a year from then (on average: 50±9 weeks). For the 13 children receiving occlusion after BBV, the AE acuity changed from 0.58±0.30 logMAR, at BBV completion, to 0.44±0.34 logMAR following an additional average period of 50±7 weeks. Whereas for 3/16, not patched, acuity remained stable for one child and showed some recurrence for the other two (both showing -0.16logMAR in their AE acuity after cessation of BBV). Overall, the mean additional gain among 16/22 was 0.12 logMAR (σ=0.24), i.e. not substantial (<0.2 logMAR) and just around the minimum change required to exceed test-retest variability for acuity (estimated between 0.1 - J. M. Holmes et al., 2001; and 0.15 logMAR - Tsirlin et al., 2015). Statistically, the mean acuity-improvement (from baseline) measured at treatment completion compared to ~1 year following BBV cessation was only marginally significant (one-sample paired t-test: $t\text{-stat}_{d.f.:15}$ =1.92, *p*=0.074). Figure 27 shows the mean *gain* in the AE acuity across children who were treated for either 8, 16 or 24 weeks, as measured after, respectively, 46 (±10), 52(±4) or 55(±30) weeks from BBV completion. Respectively, at this time-point, the mean gain from baseline was 0.35±0.17, 0.69±0.39 or 0.65±0.24 logMAR. This corresponded to a mean additional gain, from treatment completion, of: 0.03±0.20 logMAR for 9/12 children completing BBV in 8 weeks, 0.45±0.26 logMAR for the 3/4 children under BBV for 16 weeks or 0.07±0.11 for 4/6 reaching 24 weeks of BBV therapy. Paired comparisons suggest an influence of the shortest duration of treatment in limiting the final gains in acuity, but we found no statistical significance (p-values associated to each two-sample t-test are reported in Figure 27).

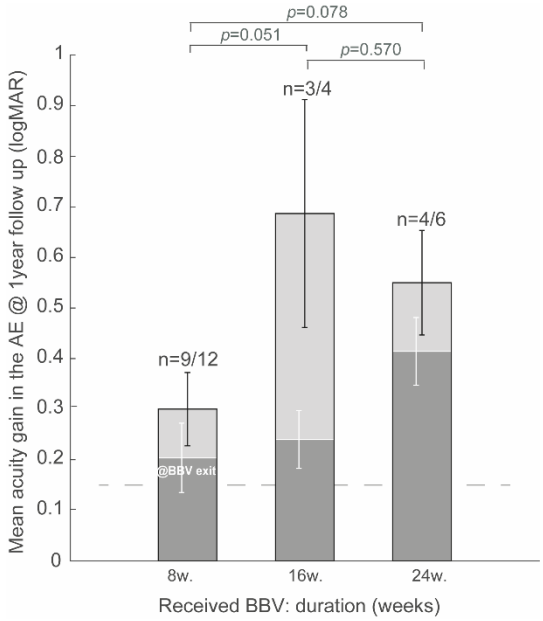


Figure 27 Mean gain in the AE acuity from baseline to ~1 additional year after treatment, obtained for children who underwent either 8, 16 or 24 weeks of BBV & attended the follow-up visit - the respective number of children for each group is indicated above each bar. Acuities were collected during standard orthoptic assessment. Dark grey bars indicate the mean gain from baseline to treatment completion, while the stacked light grey bars indicate the additional gain from completion to ~1 year after. White and dark error bars represent the standard error of the individual gains respectively obtained at treatment completion or after ~a year, for the 9, 3 or 4 children receiving BBV respectively for 8, 16 or 24 weeks.

Seven children attended a follow up at two years (mean 95 ±30 weeks after stopping BBV), with +0.01 ±0.23 logMAR mean change in acuity logMAR from BBV completion (4/7 gained 0.15 ±0.1 logMAR; 3/7 regressed 0.23 ±0.16 logMAR). At this point in time, continuing patching revealed to secure no-further improvement, either from BBV cessation (minor loss: +0.02logMAR, σ0.26) or from the 1year-follow up (minor loss: +0.05logMAR, σ0.03).

3.4 Discussion

We describe a novel treatment for amblyopia, “Balanced Binocular Viewing” (BBV) therapy, which matches stimulation across the eyes and consists of 1hr/day viewing *binocularly balanced movies* at home through shutter glasses. The procedure involves the blurring of the image received by the fellow eye (FE), at a level such that monocular FE acuity for the blurred stimulus was equal to monocular acuity for the AE (amblyopic-eye; for an unfiltered stimulus as measured at treatment induction). Twenty-two children (3-11 years), with anisometropic, strabismic or combined amblyopia, completed our study spending on average 75hrs 14min on treatment, which lead to a mean gain in acuity in the AE of 0.27 ±0.22 logMAR. This gain indicates a substantial improvement of

amblyopia (i.e. gain in the AE acuity>0.2 logMAR; Repka et al., 2003; Stewart et al., 2011), and is no inferior to those reported using alternative binocular treatments (Table 6). In terms of rate of improvement, while patching requires 120 hours of treatment for every 1 line of logMAR acuity gained (Stewart, Moseley, Stephens et al., 2004), our approach is more than four times faster yielding a similar benefit in only 28 hours of movie viewing.

In terms of stereoacuity, we obtained reliable measures (pre- & post-treatment) only in children with pure-anisometropic amblyopia - who frequently maintain a degree of binocularity, especially at low spatial frequencies (Levi et al., 2011), which remain visible to both eyes. For these children, stereoacuity reached normal values (Anketell et al., 2013) in all but one child, exceeding reports for occlusion (D. K. Wallace et al., 2011) and other binocular therapies (see Table 6).

We found that the gain in vision achieved under BBV related only weakly to reduced suppression. We note that the absence of a significant relationship between suppression-reduction and acuity-gains greatly limits the risk of inducing intractable diplopia as an adverse side effect from a binocular therapy.

We designed the study so that the treatment duration varied (i) across children, depending on the responsiveness of each child, as evaluated during orthoptic assessment in clinic and (ii) between groups of amblyopia-type, to address concerns about inducing diplopia. A pilot group of children (Group 1) was treated for 8 weeks only (according to the standard interval adopted in clinic to evaluate a visual treatment). While acuity improved in 6/7 of these children (mean gain: 0.30 ±0.26 logMAR; ID8 stable), we do not report a significant change in suppression. In the absence of any adverse events we went on to apply BBV to children in Group 2, with strabismic (N=6) and combined mechanism amblyopia (N=9), looking at longer-term effects after up to 24 weeks on treatment.

This difference in treatment duration limits the validity of comparing responsivity across children, especially since gains had not plateaued in at least some children who were released from BBV. However, we can still consider the effect of treatment length and compliance on the therapeutic outcome. Standard occlusion therapies, which are likely to improve acuity in 50-85% of children (E. E. Birch, 2013), are fundamentally limited by levels of *concordance* (i.e. agreement on treatment regimen between patient and clinician) falling below 50% (Stewart, Moseley, Stephens, et al., 2004), and *compliance*

especially at a young age (<50% in 4-7 year old children; M. P. Wallace et al., 2013). Recent studies of various binocular treatments have shown that children whose compliance is less than 50% either showed significantly lower gain in logMAR acuity (E. E. Birch et al., 2015) or required longer treatment durations to reach comparable outcomes (in 13-50yrs old participants; R. F. Hess, Babu et al., 2014). In older children, the success of monocular training (in addition to patching) was related to the total amount of training (across sessions), provided a minimum daily compliance level of 15 minutes practice per session was met (Zahra Hussain et al., 2014). In our study, on average, children who spent 8 or 16 weeks on BBV showed high levels of compliance (87 ±21 and 84 ±28%), with similar gains in AE acuity (0.2 ±0.24 and 0.24 ±0.11 logMAR, respectively). Those on treatment for 24 weeks received 98 ±46% of the prescribed dose and gained 0.41 ±0.16 logMAR (see also section 3.3.4). Interestingly, those children whose BBV treatment lasted longer spent a higher proportion of days on treatment (66, 70 and 71 % respectively over 8, 16, and 24 weeks). These results suggest that high levels of compliance, as both *daily* and *total perseverance,* can produce positive treatment outcomes (although such correlations did not reach significance in our study, perhaps due to the generally high levels of compliance). More generally, alternative treatments (to replace or augment occlusion) engage higher levels of compliance than occlusion alone (C. M. Suttle, 2010). Leaving open the question of whether occlusion is necessary, these studies highlight the importance of compliance in amblyopia treatment. Given the lack of clarity regarding the mechanism supporting improvements in vision, further study is required to determine the extent to which compliance contributes to the superior (more rapid) therapeutic response from binocular therapies.

Although it is widely assumed that the likelihood of positive treatment outcomes decreases with age (Scheiman et al., 2005), recent interventions have highlighted the possibility of improving vision in amblyopia at almost any age (Levi, 1994). As highlighted in section 3.1, 'Perceptual Learning' (PL) approaches (involving monocular training on gameplay (Bavelier et al., 2010; Zahra Hussain et al., 2014) or psychophysical tasks (Levi & Li, 2009; Polat et al., 2004) have proven effective in this regard. Gains obtained on the trained task generally transfer to acuity - showing ~1-2 logMAR lines improvement, after ~30-50hrs training depending on the severity of amblyopia (Levi & Li, 2009). Dichoptic PL has also been found to improve stereoacuity (Vedamurthy, Nahum, Huang, et al., 2015). Similar improvements have been found with binocular interventions such as, "Anti-suppression" (AST; R. F. Hess et al., 2011) or Interactive-Binocular (I-Bit; Nicola Herbison et al., 2016) therapies that seek to reduce a notional suppressive drive from the FE by equating the visibility of stimuli across the eyes.

However related studies, including games therapy in children (S. L. Li et al., 2014) and adults (R. F. Hess, Babu, et al., 2014; R. F. Hess et al., 2010b) did not exclude participants who had previously undergone occlusion therapy, making it impossible to disentangle the influence of previous treatments on outcome. Further, only a few previous studies checked for stable vision prior to treatment induction (E. E. Birch et al., 2015; S. L. Li et al., 2014). This leaves open the possibility that at least some of the therapeutic benefits of the treatment originate from ongoing benefits of optical treatment. Note that we only included children with no history of occlusion therapy and whose acuity stabilised after minimum 16 weeks of optical treatment.

In terms of stereoacuity, children with measurable (generally poor) stereoacuity at entry showed significant improvements following BBV treatment. This is consistent with earlier work - with occlusion (Stewart et al., 2005) or alternative treatment (Tsirlin et al., 2015), which indicated that the initial level of vision may limit treatment outcomes. Among perceptual learning approaches, there is evidence for a small advantage of dichoptic game-play over monocular movie-viewing in improving stereoacuity (Vedamurthy, Nahum, Huang, et al., 2015), although monocular game-play can also be effective (Levi & Li, 2009). Anti-suppression treatment has been shown to significantly improve stereoacuity in adults (R. F. Hess et al., 2010a; R. F. Hess et al., 2012; To et al., 2011) but not always in children (E. E. Birch et al., 2015; S. L. Li et al., 2014). Further research (e.g. within a randomised controlled trials) is needed to determine which components of these therapies are critical for triggering improvements in specific visual functions (such as stereoacuity) and to establish the wider applicability of these treatments (Vijay Tailor, Bossi et al., 2014).

To equalise acuity across the two eyes, we individualized the level of Gaussian blur applied to the image viewed by the FE during movie-viewing (i.e. the high SFs were attenuated in proportion to the AE acuity deficit). This is in contrast to the fixed blur levels used in treatments that rely on e.g. Bargerter translucent filters (Rutstein et al., 2010). In contrast, Anti-suppression therapies manipulate the *contrast* of the signal to balance visibility across the eyes, and update this level as the child's vision changes during therapy. The fact that we observe substantial gains in acuity indicates that fixed levels of blur penalisation and contrast penalisation are both effective for treatment.

Whether and how binocular treatment methods are related to suppression in amblyopia is hotly debated. Some hold that suppression is a *cause* of amblyopia (R. F. Hess, Thompson, & Baker, 2014; J. Li et al., 2011) and that Anti-suppression therapies

strengthen binocular combination by breaking this suppression, allowing monocular and binocular vision to improve. If this were the case, we would expect better outcomes (i.e. improved acuity) to be associated with greater reductions in suppression. We, like an earlier study (Vedamurthy, Nahum, Bavelier, et al., 2015), do not find strong evidence for such an association, although we note that our approach did not produce substantial changes in suppression at all, unlike at least some binocular gaming therapies (Vedamurthy, Nahum, Bavelier, et al., 2015) and occlusion therapies (Kehrein et al., 2016). That BBV treatment is effective in the absence of changes in suppression is however consistent with the idea that suppression cannot be the sole cause of amblyopic visual loss. An alternative view is that suppression is a *consequence* of amblyopia in order to avoid diplopia (Holopigian et al., 1986). If so, a temporary disruption of binocularity (e.g. using rTMS) should have no impact on monocular function. That it does (R. F. Hess & Thompson, 2013) supports the presence of reduced (not lost) functionality of the AE, possibly as a result of suppression.

If gains in acuity do not originate from a reduction in suppression, what does produce them? Current theories of the neural basis of amblyopia focus on the consequence of abnormal input from the AE for neural encoding within the LGN and V1. Candidate models are: a reduction in the number and/or sensitivity of neurons driven by the AE - *under-sampling* (R. F. Hess & Anderson, 1993; Levi et al., 1999; Sharma, Levi, & Klein, 2000), positional disorganisation of visual receptive fields and associated distortions of retinotopic mapping - *disarray* (Clavagnier et al., 2015; R. F. Hess & Field, 1994; Zahra Hussain, Svensson et al., 2015), and increased variability in the response of binocular cortical neurons - *elevated noise* (Levi, Klein, & Chen, 2008). Animal models of amblyopia have produced results to support each of these mechanisms (Sengpiel et al., 1998), though the magnitude of these deficits in V1 rarely matches the scale of the behavioural deficits, suggesting an additional role for brain areas beyond V1 (Kiorpes et al., 1998). In human vision, Clavagnier *et al* (2015) recently reported findings from fMRI population receptive field (pRF) mapping: they reported an *in vivo* estimation of visual receptive field size and density (Dumoulin & Wandell, 2008; Harvey & Dumoulin, 2011) from the FE and AE of patients with amblyopia. They observe larger pRFs in the fovea of amblyopia patients across areas V1-V3, but normal cortical magnification across these areas, which could arise either through under-sampling, positional disarray, or both. This fits with the more general finding of increased receptive field sizes in binocular V1 neurons following retinal lesions (Gilbert & Wiesel, 1992). Given the dependency of the size of RFs (and by extension pRFs) on visual experience in these studies, we consider the *reduction* in RF size to be the most reasonable current candidate for the

mechanism producing therapeutic response. We will call it the *'re-mapping'* mechanism. This in turn could be driven by reductions in under-sampling and/or disarray within a range of cortical regions, as above. Importantly, the dissociation between suppression and acuity gains in our study suggests that although suppression may play a causal role in these amblyopic deficits (J. Li et al., 2011), the mechanism underlying such deficits can be altered by treatment without concomitant changes in binocularity, i.e. without modifying suppression.

3.4.1 Conclusion

Our Balanced Binocular Viewing treatment engages high levels of compliance and leads to substantial gains in visual function after a relatively short period of treatment. BBV is an engaging and unsupervised binocular vision treatment that also supports remote monitoring of compliance and suppression. Our findings thus far indicate that a reduction in inter-ocular suppression is not the basis of the observed improvements in visual acuity.

	Occlusion (patch or atropine drops)	Game play & **Perceptual Learning**	Game play: AST **(anti-suppression** therapies)	I-Bit (**Interactive** Binocular treatment)	BBV (**Balanced** Binocular Viewing)
Methodology	Enhanced usage of AE	Repetitive gameplay and psychophysical experiments either with occlusion of FE* or not**	Video games with elements distributed across eyes. Contrast imbalance is progressively reduced.	Modified video and interactive games (only AE sees key details of the scene)	Movie viewing in balanced stereoscopic presentation (FE vision blurred)
Main published **studies**	(T. Li & Shotton, 2009; Repka et al., 2005; Stewart et al., 2011) ●RCT: -ROTAS gr- (Stewart et al., 2007) patching for 18w ODM: 6vs 12h/d -PEDIG gr-1: (J. M. Holmes et al., 2016) 2h/d patching vs binoc. games for 16w(7d/w) -PEDIG gr.-2: **OT** vs binoc game (ongoing) -RODS (protocol RCT): personalised vs standard dose occlusion (Moseley, Wallace et al., 2015)	* (Zahra Hussain et al., 2014; for PL review: Levi & Li, 2009; Polat et al., 2004; Xi et al., 2014) ** (Vedamurthy, Nahum, Huang, et al., 2015) ● phase1 RCT: NCT01223716- PL vs games vs ~2h/d patching- for 1to6m.(5d/w)	(R. F. Hess et al., 2012; Knox et al., 2012; S. L. Li et al., 2014; To et al., 2011) ●RCT: i) (T. Y. Gao et al., 2018)vs placebo games; ii)(K. R. Kelly et al., 2016)vs 2h/d patching for2w.(7d/w)	●RCT: (Nicola Herbison et al., 2016)	(Bossi, Tailor et al., 2017)
Mean VA gain in logMAR units (in diff. studies)	~0.15-0.3 ROTAS:>3h/d is more effective; PEDIG:0.14 (vs 0.10)	~0.2	0.14 (0.08-0.19) i)0.06(vs0.07) ii)0.15(vs0.07)	0.08 (video:0.1-games:0.06)	0.27
Compliance (dose received vs. prescribed)	44% (M. P. Wallace et al., 2013)	(~50-100%)	(100%)	>90% (exactly n.s.; supervised)	90% (remotely supervised)
Recurrence (following loss of VA gain)	~30% at 1 year (Bhola et al., 2006; J. Holmes et al., 2004)	Small decrements to nil.	n.s.	At 10w.: Video-0.03/ games-nil	At ~50w. (n=16): 0.07
Age ('best-fit')	Pre-s. children	Adults	Adolescent/adults	Pre-s./school children	Pre-s./school children
Treatment **duration**	~12-24 w. (2to12h/day)	Up to 39w. (PL:6-50h, up to 522h)	1-9 w. (0.5-2h/sessions)	6 w. (30min/w.)	8 w.-24 w. (1h/day)
Setting (**supervision** required?)	Clinical (yes)	Clinical (and home (Zahra Hussain et al., 2014; Vedamurthy, Nahum, Huang, et al., 2015) (yes)	Clinical (home(S. L. Li et al., 2014)) (recommended)	Clinical (yes)	Home (recommended)

Table 6 An indicative summary of current and proposed treatments for amblyopia. Main RCT (randomised control trials) studies are indicated. Note: data are pooled from representative studies (most cited in this chapter); the table is not a complete review of the current literature. Column 1 describes occlusion therapy (i.e. current clinical practice; ODM=occlusion dose monitor; OT=optical treatment), columns 2 shows approaches that supplement occlusion, and columns 4-6 list alternatives to occlusion (i.e. binocular treatments). Key: n.s.=Not Specified; VA=visual acuity; AE=amblyopic eye; FE=fellow eye; m.=month(s); w.=week(s). The listed research groups are: (UK based) MOTAS=monitored occlusion treatment of amblyopia, ROTAS=randomised [~], RODS=randomised occlusion dosing strategy. (US based) PEDIG=Paediatric eye disease investigation group (running various ATS=Amblyopia Treatment Studies).

4 Effect of BBV treatment on crowding and contrast sensitivity

Does BBV treatment affect contrast-sensitivity? And crowded-acuity? If so, do the children included in our study manifest a pattern of deficit in line with what other studies have reported?

4.1 Introduction

In the previous chapter we described an exploratory study of a new treatment for childhood amblyopia (BBV). We included children with anisometropic, strabismic and combined mechanism amblyopia. As described in Chapter 2, in these types of amblyopia unilateral image degradation occurs caused by either a difference in refractive power, a misalignment of the two eyes or by a combination of these factors. This results in a syndrome of functional visual deficits. In case of pure anisometropia, it is commonly assumed that each dioptre of interocular difference induces about 1% reciprocal difference in the retinal-images' size (tolerated until ~5% difference; Achiron et al., 1997), with the consequent typical 'blurred vision' through the weak (amblyopic) eye. This is associated with abnormal contrast sensitivity in this eye, over a full range of spatial-frequencies, with higher vulnerability at high SFs (Bradley & Freeman, 1981; Zele, Pokorny et al., 2007). However, binocularity tends to be maintained with some degree of stereoacuity usually being measurable. Further, spatial acuity is only moderately affected (Agrawal et al., 2006). In contrast, strabismus, or squint, either manifest or latent, leads to fixation instability which disrupts binocular vision. Strabismus is thought to spare monocular contrast-sensitivity to low SFs, as each eye, independently, would not be 'disturbed', whereas sensitivity to high-SFs would be selectively affected, proportionally to the reduced acuity of each eye (McKee et al., 2003). Therefore, foveal crowding might occur more markedly in presence of strabismus, as stimulus configuration at high SFs intrinsically contains more repetitive periodic patterns (Levi & Klein, 1985); while the effect of a cluttered presentation would be limited in anisometropia (Greenwood et al., 2012), when a perspective from both eyes can be compared.

We developed a series of psychophysical measures to investigate the pattern of visual loss in amblyopia, and how it changes following treatment. Data were acquired on the same children who participated in our BBV study (discussed in Chapter 3). Specifically, we used 4 tests (acuity, contrast-sensitivity, crowded-acuity and stereoacuity) to collect additional evidence for evaluating the effect of BBV treatment, pre- versus post- home-based therapy. Full details of this psychophysical testing-procedure are given in a published study (Greenwood et al., 2012).

4.2 Methods

4.2.1 Participants and procedure

Before commencing, and after being released from, BBV treatment, the experimenter collected a series of behavioural data, to complement the results obtained during the planned orthoptic assessments (at 4,8 and possibly 16 and 24 weeks after BBV therapy was commenced; see section 3.2.3).

Behavioural testing took about 45 minutes per child. Details of the procedure are summarised below. The tests were administered in the following order: acuity (necessary for the tests that followed), crowding and contrast, or vice versa - depending on the need to maintain the attention of the child, and finally, stereoacuity (most likely to be unmeasurable in children with strabismic amblyopia). The experimenter spread testing over multiple sessions/days when the child was inattentive. Breaks were given, including e.g. snacks, drawings etc. to sustain the child's interest. When a second home-appointment was necessary, were the child to still not comply, BBV treatment was commenced anyway. The two additional behavioural tests, necessary to set the level of blur applied to movie-images for the duration of the treatment (see section 3.2.3 for details) were necessarily prioritised. Following successful completion of these tasks, tests on AE vision were prioritised. On two occasions, an unforeseen change in a family's schedule led to limited data collection on the day it was agreed. The experimenter offered to attend for a new visit within a week. When this was not possible (1/2 cases), treatment was commenced and data collection forgone. To collect post-treatment data, as many tests as possible were administered on a single day, when BBV-system's recollection had been agreed with the patient and parents. In exceptional circumstances, and when the child's entry data had not been missed (alternatively, the matching post-treatment data were not measured), the opportunity for a second appointment was offered. For 3/22 children a second visit was necessary, mainly due to the lack of attention of each child.

4.2.2 Behavioural tests

The tests are fully described elsewhere (Greenwood et al., 2012). In short, the battery includes four behavioural tests (visual acuity, contrast sensitivity, stereo-acuity and crowding), each taking approximately 15 minutes to administer. During all tasks, the child wears stereo-shutter-glasses allowing us to control what each eye sees. A character, called Vac-Man ("Visual acuity man), is presented in the centre of the screen surrounded by four flanking "ghosts" (visible through both eyes; an example in Figure 28). The child has to discriminate the direction Vac-Man's mouth is pointing. This can be done verbally ("up", "down", "left" or "right") or by indicating the ghost (by position or colour) that VacMan is about to "eat". In the stereoacuity test, the stimuli are formed by four random-check stereograms modulated: the child has to indicate the "odd-man-out", that appears either in front of or behind the plane of the display. Stimuli remained on the screen until the child responded at which time the experimenter recorded their response using a keypad.

All tasks, except stereoacuity, tested the monocular performance, obtained independently under FE-view and AE-view, in this order. Entertaining animations followed every 3 consecutive correct responses: here Vac-Man was shown smiling and moving toward the 'ghost' he wanted to eat. Following each trial, a smiling (correct) or sad (incorrect) face provided feedback. Animations were accompanied by engaging sounds to encourage participation.

A QUEST adaptive staircase procedure (A. B. Watson & D. G. Pelli, 1983) estimated the individual threshold for each task by changing either (a) size, in the acuity task, (b) contrast, in the contrast sensitivity task, (c) distance from the flanking ghosts, in the crowding task and (d) offsets of dichoptic ghost-images (modulated in depth) in the stereoacuity task.

In the **acuity** discrimination task (always performed first) a dark Vac-Man was centrally displayed at 50% Weber contrast against a grey background. Target-size (starting at 6 deg, similar to the size of a 0.6 logMAR target at 3m) was controlled by a QUEST adaptive-staircase procedure, converging on 62.5% correct performance. We included catch trials every five stimulus-presentations (displaying stimuli at twice the current threshold estimate) to maintain motivation of the participant and to remind them of the appearance of the target. After three practice trials, QUEST ran for a maximum of 32 trials unless σ of precedent eight trials was less than 0.03 log units (Greenwood et al., 2012), in which case the threshold was set in advance.

The **contrast detection** task (performed second or third) was similar, except that the staircase procedure modulated the target contrast (starting at 50%) instead of size. Stimuli were presented at sizes equal to 2.5×acuity thresholds for each eye, to ensure that visibility of the maximally-cued (contrast or crowded) target was matched across eye and observer's sensitivity.

The **crowding** discrimination task (performed second or third) used similar target-scaling as the preceding task but differed from other tasks in that ghosts were achromatic/dark (50% contrast), had no eyes and were presented at random (N, S, E,W) orientations. In this way, target and ghosts were similar enough to induce crowding. The QUEST procedure now altered the centre-to-centre separation of target and ghosts, starting at six times the Vac-Man radius, to estimate the minimum separation allowing identification of the target-direction. When target and flankers abutted, the procedure continued for three trials and exited if no errors were committed (Greenwood et al., 2012) - otherwise maximum 32 trials were presented.

In the final test of **stereoacuity**, the four ghosts were formed from a random-checkerboard carrier and a "ghost-shaped" disparity-defined region defined. The target was the only ghost presented with some binocular disparity, while the others lay in the zero-disparity plane: children were asked to identify the ghost that popped out the screen. Random checks were light or dark at 50% Weber contrast and corresponded at 1 monitor pixel (if not enough we increased the dimension at 2, 4 or 8 pixels per random-check). Three practice trials (or more if required) anticipated the beginning of QUEST, starting at 160 sec of arc disparity (~0.04 degrees) and then varying the binocular disparity of the target-ghost.

Consent and assent taken prior to inclusion in BBV study also covered the pre- and post-behavioural assessment of vision using the approved computer system and 3D goggles to control monocular or dichoptic presentation of the stimuli (Figure 28). The battery was approved by local and National Health Service ethics, and conformed to the Declaration of Helsinki.

Figure 28 Schematic representation of “VacMan” test. The child responded (either verbally or by pointing) which “ghost” was about to “eat”. Children always wore LCD shutter glasses to allow us to control the stimulus seen by each eye. The experimenter entered the child’s response using a keypad

4.3 Results

All these behavioural data were acquired using our computer-tests, under monocular viewing through the AE, while the FE was occluded (by mean of 3D shutter glasses). Data-sets for the three monocular tests (acuity, contrast and crowding tests) were *incomplete* for 12/22 children. Among them, ID4 did not complete only the FE-post-treatment acuity test. Note that FE-tests were the first to be omitted when shortening the testing-session. Instead, for 3 children we missed at least one measurement relative to the AE for each test. In one of these cases time was the limiting factor. For the other 2 children, we considered their young age, and decided not to insist. In particular, based on behaviour and attitude during the pre-treatment testing, we anticipated that one child (ID17) would not be compliant with treatment (although consent had been taken at this stage). Indeed, this child received the lowest daily dose among all participants (on average, 42% i.e.25 minutes per day, on 42% of days when the system was used, on average). We note however, that this level would still be considered ‘good’ on a scale usually applied to compliance-to-patching, as ID17 exceeded 30% of the 60 minutes prescribed daily.

Focusing on the AE, the diagram in Figure 29 indicates for each child, and with respect to each test, which AE-data we missed. So, AE-acuity data-set was incomplete for 4/22 children (only ID17 missed both pre- and post-treatment AE-tests). Contrast data were the least complete among the three tests, with 9/22 children being unwilling to complete the session (when using either eye), plus one (ID16) whose performance was classed as an outlier (not represented in Figure 30c). Finally, for 7/22 children we did not get measurable performance on the crowding-test (3/7 for both eyes).

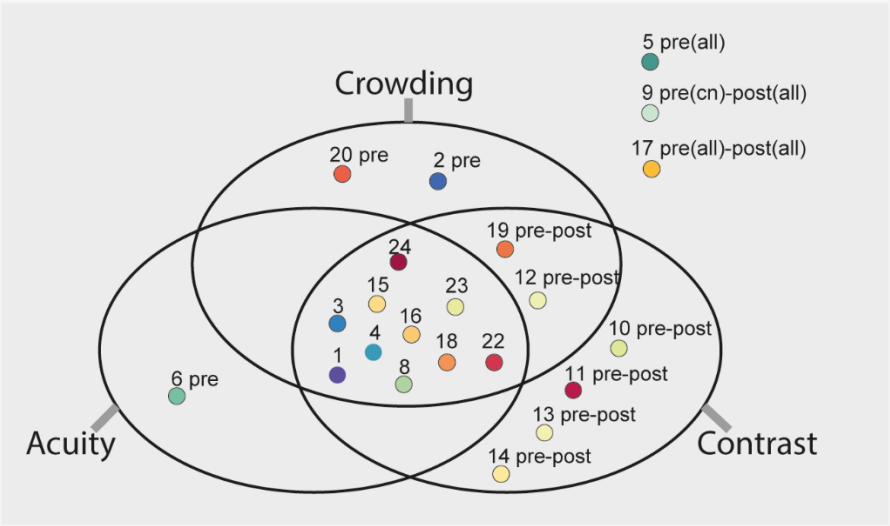


Figure 29 Psychophysical data-sets. We report for each child (coloured dots, labelled with the child's ID for the BBV study- details in 3.2.1) which measurements relative to the AE were *incomplete,* for each test's data set (three sets, labelled; cn=contrast). Overall, data sets were incomplete for 12/22 children (with IDs 5,9,17 who missed at least one AE-measurement per test).

Data set for stereoacuity have not been included in the analysis as only 2 children (with anisometropic amblyopia) succeeded in completing both the pre- and post-treatment measurements. For the remaining anisometropes (6/8 children), QUEST procedure did not converge to threshold, despite at least four attempted measurements per child (two for each eye). Among the included children diagnosed with strabismic (N=6) or combined mechanism amblyopia (N=9) we did not proceed beyond ~5 stimuli presentations, as responses were clearly at random and we were aware of their unmeasurable stereopsis using clinical tests. An exception was made for four children (ID20, 18, 15 and 14), who probably guessed right a few trials; but they also did not reach a threshold.

Figure 30 compares (a) acuity, (b) contrast sensitivity and (c) crowding (quantified as the minimum target-flankers' separation allowing for direction-detection of the target's aperture) measured pre- and post-treatment (respectively, during system installation and recollection). Data points in the three panels are colour coded in ascending order of individual IDs (numbered 1 to 24; IDs 7 and 21 omitted: these two children did not complete BBV treatment). This colour-coding also gives us a visual indication of children belonging to Group 1 or Group 2, i.e. Gr.1: purple-blue symbols are children with pure anisometropic amblyopia, IDs 1-8, treated for maximum 8 weeks. Gr.2: green/yellow-red symbols: children with strabismic and combined-mechanism amblyopia, IDs 9-24, treated for up to 24 weeks. We notice that the majority of the data lie around or below the line of equality (x=y), indicating respectively a non-negative or positive response to treatment, reflected in reduced thresholds post- (vs. pre-) treatment.

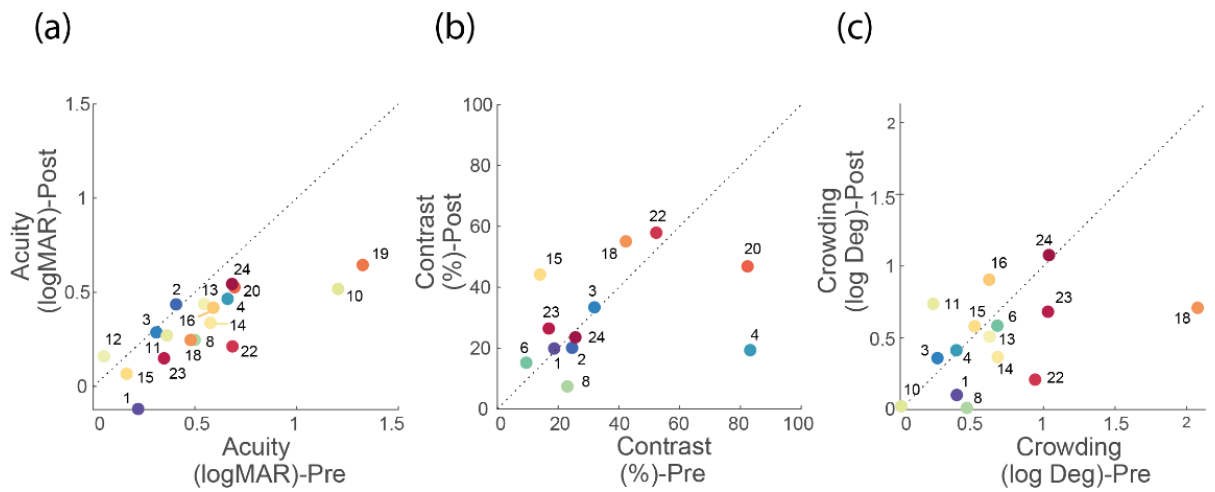


Figure 30 A comparison of pre- versus post-treatment psychophysical results, respectively to acuity (a), contrast (b) and crowding (c) tests, administered using the same PC-screen adopted for BBV treatment, under monocular AE view, while occluding FE. Each data point is labelled with the ID corresponding to the child completing the test. Colours identify the order of recruitment: IDs 1-8, children with anisometropic amblyopia (purple to blue hues), Gr. 1; IDs 9-24 children with strabismic and combined mechanism amblyopia (green/yellow to red hues). A dashed line in each plot represents the line of equality: data points below the line indicate an improvement in the measured visual ability.

A one-sample paired t-test revealed a significant change in acuity (t=4.26; p=0.0006), in line with the results obtained using logMAR crowded charts in the clinic. Among 17/22 children (having a complete data-set), acuity improved 0.23 logMAR (σ: 0.21) on average. The disturbing effect of crowding was not significantly reduced (t=1.25; p=0.2227), as well as contrast sensitivity did not significantly improve (t=0.71; p=0.4940).

To check whether greater changes in contrast sensitivity and in susceptibility to crowding were associated with greater changes in acuity we looked at paired correlations between the changes in thresholds acquired pre- and post- treatment across different tests. Figure 31 (left) suggests a moderate correlation between improvement in acuity and reduction of crowding effect, but this does not achieve significance (R=0.22, p=0.46; ID10 and 18 were excluded from the correlation analysis as outliers). The correlation between individual variation in acuity-change and the change in contrast sensitivity is clearly non-significant (Figure 31-right; R=0.04, p=0.91; ID16 excluded from the correlation analysis as outlier).

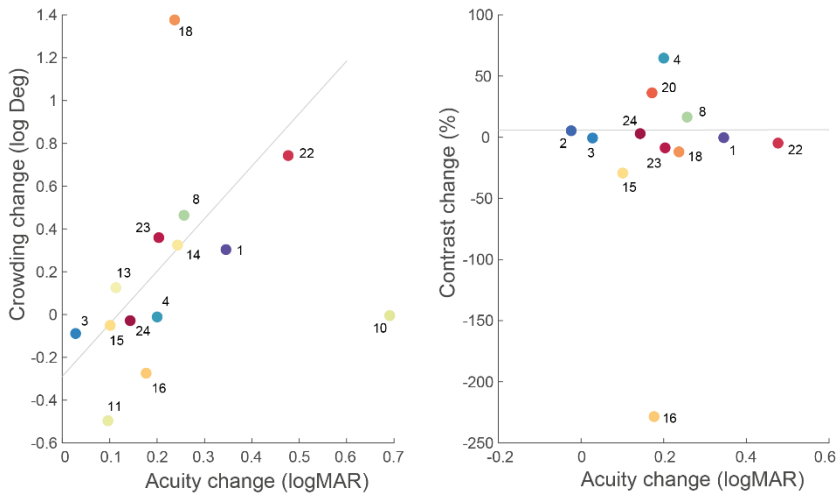


Figure 31 Acuity, contrast, crowding. A comparison between change in acuity and (left) change in crowding or (right) change in contrast sensitivity. Data points are colour coded as in Figure 30. On the left, IDs10 and 28 have been labelled outliers and excluded from the correlation analysis; on the right, the same applies to ID16. Neither of the two correlations were significant at α=0.05

4.4 Discussion

Summary of results. The positive therapeutic effect on acuity of our BBV treatment (described in Chapter 3) is confirmed by the results obtained psychophysically, although the amount of gain doesn't match. However, a mean mismatch of 0.05 logMAR is not surprising considering the different tests respectively used during the clinical and the experimental assessments. We also noted that in both pre- to post- assessments (clinical and experimental), the minimum difference required to detect a reliable change in acuity (~ 0.15 logMAR in children; for a review see Anstice & Thompson, 2014) has been exceeded. A justification for this offset in acuity-mean gain might be that we measured single letter acuity, subject to over-estimations, while the clinical assessment used crowded charts. In fact, excluding those children for whom we could not measure acuity psychophysically, we found a mean acuity in the AE equal to 0.78±0.34 logMAR, using crowded charts in the clinics, or equal to 0.57±0.32 logMAR, using our computer game-test. In support to the mentioned justification, we reported a tendency (although not significant) towards a reduction in the susceptibility to crowding being more marked the greater was the gain in acuity – as shown in Figure 31b, this was more evident in some children whose initial crowding thresholds were visibly higher than those scored post-treatment (e.g. ID 1, 8, 22 and 23). Unfortunately, stereoacuity was not measurable in the vast majority of children, therefore we cannot consider the relationship between acuity, stereoacuity and crowding (also when considering the clinical Frisby results, the data set is incomplete to allow for e.g. ANOVA).

Note that by using our game-test, we had less ‘space for post-treatment improvement’ than if using crowded charts. However, on top of having a comparative measure to the clinical results for acuity, it was important to let the child familiarise with the simple but fundamental tests needed to set the level of blur to be applied to the FE-image throughout treatment (see 3.2.3). This is why, here, we used this type of acuity test.

Incomplete data set. We obtained a complete data set for the three monocular tests for 10/22 children included in the BBV study; while 12/22 did not. The main difficulty was to maintain the attention of each child for the entire duration of both pre- and post- testing, respectively planned on the same day of installation and of recollection of BBV system. In particular, during installation (anticipating how the testing was going to be conducted during recollection), the experimenter had to set the system, answer any question and, importantly, dedicate enough time to the testing and set-up of the individual level of blur (see 3.2.3 for more details on the BBV regimen). Remember that children wore their prescription-glasses under the shutter-goggles. However, we questioned whether a poorer level of initial vision in the 12/22 children who did not complete the tests, compared to those who did, could have caused a faster loss of attention, maybe due to a distress induced by the blur applied. Nor the baseline level of acuity, nor the initial IOAD were significantly different between children who did- or did-not complete the tests (respectively, 2 samples t-test: p=0.533, 0.748). Of these 12/22 children, 3 had anisometropic amblyopia (43% of the included children with this type of amblyopia), 2 strabismic (33%) and 7 had combined mechanisms amblyopia (63%). Therefore, the type of amblyopia was not informative in this sense. Also, their respective age was only marginally different (2 samples t-test: p=0.09) and we noticed that only 4 of the 10/22 children younger than 5 yrs were among those not willing to complete the tests. It was parents’ decision when to fix the home-appointment. The experimenter knew only in some occasions what other activity, of his/her routine, the child was ‘replacing’ while completing the tests. It could be that having to miss a ‘fun’ activity (e.g. playground) had a particularly negative impact on attention. But this, and the supportive attitude of parents, were variable out of the experimenter’s control.

Acuity and contrast testing. We measured monocular acuity as the ability to discriminate exclusively high contrast details (i.e. fine spatial scale; in this case, the mouth’s direction of the central character). Further investigation could address binocular acuity, as spatial-scale has been demonstrated to affect binocular combination in amblyopes, in which fusion is less likely to occur when discriminating fine scaled dichoptic stimuli (Spiegel, Baldwin, & Hess, 2016). However, if any difference in binocular acuity was present, we

should have found evidence of change in interocular suppression measured respectively on the same participants, as strong suppression prevents fusion and therefore binocular acuity. The fact that we did not show significant changes in suppression allows us to hypothesise that binocular acuity measures would not add much to the analysis.

Improvements in contrast sensitivity might have a prognostic value for treatment-outcome in visual acuity, as the gain in acuity relates to contrast-sensitivity change, at least in term of reduced interocular contrast-balance (see for example Eileen E. Birch, Morale et al., 2016; refer also to 2.8.2 ; Ding & Levi, 2014; R. F. Hess & Thompson, 2015). Weather this is true in our sample of children cannot be confirmed by referring to the series of experimental-tests presented in this chapter: for example, IDs 4, 8 and 20 all showed a greater change in contrast than the one measured for other children, but their respective change in acuity was similar or lower than others' - who instead showed almost no change in contrast performance (ID1, 22, 24). Note that we measured *monocular*-contrast. Similar results were obtained in the previous study using the same testing-procedure (Greenwood et al., 2012). Therefore, the testing method might have a huge impact on the relationship between contrast- and acuity- results; attention should be given in particular to the relation between acuity and monocular vs. interocular-contrast.

Implications and further research. There is increasing interest in developing an accurate and practical test to measure contrast sensitivity in the clinic that would contrast the either inefficient or highly specialized (and expensive) testing equipment currently available and that could integrate newly acquired knowledge (Pelli & Bex, 2013), e.g. about plasticity and perceptual learning. New techniques have been proposed, that adapt classical charts (Hou, Huang et al., 2010; Thayaparan, Crossland, & Rubin, 2007) and eventually implement them on digital devices (Dorr, Lesmes et al., 2013). Similarly, there is increasing interest around crowing-effect, especially since a) crowded-acuity letter charts are widely available and markedly more sensitive than single-letter charts, for either non-amblyopic (Lalor, Formankiewicz, & Waugh, 2016) or amblyopic viewers (Anita J Simmers, Gray, & Spowart, 1997), and b) there is increasing awareness about the functional deficit in amblyopia and its relation to crowding (Greenwood et al., 2012; Norgett & Siderov, 2017). Our set of tests has the potential of being a quick and entertaining way to measure acuity, contrast, crowding and eventually stereoacuity in children; it has the advantage of using the same platform across all of these tests and, as typically all the computer-based tests, could be easily adapted e.g. to any viewing

distance during testing sessions. More investigation would be needed to test the validity and reliability of these tests compared to those available on the market.

Unfortunately, we were not able to adapt these tests to anaglyphs glasses presentation. Although the possibility to control what each eye sees (i.e. a dichoptic presentation) would have been maintained, vision through anaglyphs filters is extremely influenced by the gamma correction and uniform luminance of the display, so that each eye might receive an input that does not match in luminance and chromaticity – cross-talk is more likely to occur than using other 'segregation-techniques' (Daniel H. Baker, Kaestner, & Gouws, 2016). So, monocular testing would have been particularly difficult.
Note that we have conducted a pilot study of a different series of tests -discussed in Chapters 5 and 6, using anaglyphs instead of shutter goggles. The performance of normal adults was highly variable between sessions of the same test and results tended to show no-differences across participants; in contrast to the results obtained using shutter-goggles. Although we used different tests to those described in this chapter, acuity and contrast-sensitivity are common visual abilities required for a successful performance in both series of tests (those described in this chapter, and those presented in ch.5-6). Moreover, we would anticipate equally- if not more- *un*reliable results on children. Therefore, our initial plan of using shutters can be justified. Future advances in research and technology might help reducing the cost of shutter-system in favour of anaglyphs.

This chapter is adapted from the submitted paper: Bossi M. et al, "A comparison of tests for quantifying sensory eye dominance" submitted to *Vision Research,* Jan 2018 (manuscript VR-17-352 currently under revision); Elsevier's license permits posting to UCL repository, for non-commercial use.

5 Quantitative tests for binocular vision

In Chapters 3 and 4, we presented our BBV treatment for amblyopia. After on average 75 hours of modified dichoptic-movie viewing, 22 children experienced a mean improvement in visual acuity of almost 3 logMAR lines. Following BBV treatment visual acuity improves by 0.1 logMAR, in about 25% of the time usually required for patching. We also note that such therapeutic gain under BBV could not be solely attributable to a reduction in inter-ocular suppression, which has been the goal of alternative binocular treatments for amblyopia. Further, as discussed in Chapter 4, there is no consensus as to the best way to quantify inter-ocular suppression. We used a method using contrast-polarity rivalrous stimuli but other groups have used contrast-motion rivalrous stimuli or binocular summation of phase or shape. Thus, we cannot rule out that the task used to probe suppression influences the pattern of suppression reported (Richard Harrad, 1996), in line with previous results demonstrating different results depending on the test used to measure binocularity (Jonathan S. Pointer, 2012; Rice, Leske et al., 2008; C. Suttle, Alexander et al., 2009), including measures of stereoacuity (Heron & Lages, 2012).

We therefore next sought to compare the performance of different tests measuring 'suppression' to both evaluate whether they tap into common (e.g. suppressive) mechanisms and to determine the suitability of these tests for use in clinical practice.

5.1 Introduction

When combining signals from the two eyes into a coherent 'cyclopean' percept, observers rely (to differing extents) on one eye more than the other. Such *eye dominance* is important in a number of clinical settings. First, imbalances between the eyes in childhood can lead to severe loss of acuity in the weaker eye, a condition called amblyopia (E. E. Birch, 2013). Treating amblyopia (e.g. by patching) attempts to "rebalance" vision, by improving acuity in the weaker eye. Thus, the relative contribution of each eye to cyclopean vision is a key measure for understanding amblyopia. This is

particularly the case as there is debate as to the extent to which suppression of the weaker eye contributes to (e.g. E. E. Birch, 2013; R. F. Hess & Thompson, 2015; Wong, 2012) or results from (e.g. Bossi et al., 2017; Kehrein et al., 2016; Vedamurthy, Nahum, Bavelier, et al., 2015) the condition, or whether this is even a reasonable dichotomy. Second, presbyopia (an age-related refractive error) can be corrected using monovision, which enhances intermediate-distance vision in the dominant eye only, and so necessitates a decision on eye dominance. Finally, loss of visual function in age-related macular-disease leads to subtle changes in patients' reliance on their two eyes ("binocular balance"). A simple method for quantifying changes in sensory eye dominance may therefore have diagnostic value for this and other conditions (Wiecek, Lashkari et al., 2015).

A variety of methods are available to measure the relative contribution of each eye in a clinical setting. Clinical measures of eye dominance tend to be intuitive, for example assessing with which eye a patient can more easily wink (Miles, 1930). Building on the notion that eye dominance is linked to motor control, comparisons of eye-alignment are still routinely used. For instance, in the *Miles test* and *Porta test,* near and distant targets are aligned binocularly, and then each eye is opened and closed in succession to reveal which eye gives the more accurate estimate of position. Such tests of 'sighting' dominance are binary (left or right) and tend to elicit variable results (Johansson, Seimyr, & Pansell, 2015), that are dependent on specific test and viewing conditions (Rice et al., 2008).

Sensory eye dominance (SED), on the other hand, refers to the relative *perceptual* contribution of each eye to the cyclopean percept under dichoptic conditions (Ooi & He, 2001). In the clinic, *Worth's four dot test* produces a qualitative estimate of SED. This task has the observer wear red-green anaglyph glasses and report the perceived number and colour of 4 illuminated circles (1 red, 2 green and 1 white). The only *quantitative* direct clinical test of SED is the *neutral-density filter bar* (a version using red filters is the *Sbisa bar*), that quantifies the minimum filter-density that must be placed over one eye to induce the use of the other eye (McCormick, Bhola et al., 2002). However, such tests can be challenging for children which likely contributes to their only moderate test-retest reliability (Crawford & Griffiths, 2015; M. Piano & Newsham, 2015) and their infrequent use for paediatric screening (Vijay Tailor, Balduzzi et al., 2014).

Neither sighting-dominance nor sensory dominance assess the functional advantage of having two eyes. When each eye makes a relatively equal contribution, observers

experience binocular summation (increased contrast sensitivity when using both eyes; D. H. Baker, Meese, & Hess, 2008; D. H. Baker, Meese et al., 2007; D. H. Baker, Meese, & Summers, 2007) and stereopsis (the perception of three-dimensional structure based on retinal disparity; Cumming & DeAngelis, 2001; O'Connor et al., 2010). Although there is no standard clinical test of summation, stereoacuity is the most widely accepted clinical measure of binocular vision. Stereoacuity tests quantify the minimum retinal disparity supporting reliable discrimination of surface-depth (Julesz, 1971). Popular variants include the *TNO*, the *Randot* test (E. Birch, Williams et al., 2008) and the *Frisby* test (Frisby, Davis, & McMorrow, 1996). Stereoacuity tests are simple to administer and can easily be explained to children. However, they have drawbacks; some contain monocular cues (Fricke & Siderov, 1997), they can be difficult to administer in the presence of strabismus (McKee et al., 2003) and they are not necessarily indicative of clinical conditions (since 1-14% of the general population are stereo-blind; Bosten, Goodbourn et al., 2015) which limits their utility for vision screening (Cotter, Cyert et al., 2015). For example, despite being stereo-blind patients with strabismic amblyopia can demonstrate normal binocular summation when balancing the visibility of dichoptic stimuli (D. H. Baker, Meese, Mansouri, et al., 2007).

Outside the clinic, where stereoacuity is the foundation of binocular assessment, vision researchers have developed diverse, innovative and quantitative measures of sensory eye dominance. Rather than directly asking observers to make a judgement about contrast similarity (as required for e.g. the Sbisa bar), psychometric assessment tends to rely on forced-choice methods (for a study that linked traditional clinical methods of SED to a common psychophysical measure see J. Li, Lam et al., 2010). Pairing robust psychophysical judgements with carefully-designed stimuli allows assessment of the relative contribution of each eye to visual processing in different stages of the cortical hierarchy.

Here we consider SED tests first according to whether they do or do not preclude a coherent cyclopean percept (i.e. are intrinsically *rivalrous*). One test which poses no cyclopean conflict presents observers with dichoptic sine-wave gratings that differ only in phase and contrast (Figure 32, test #4). Observers are asked to indicate the location of the middle dark stripe of the phase-shifted grating that results from binocular summation (Ding & Sperling, 2006; C.-B. Huang, Zhou et al., 2010; C. B. Huang et al., 2009; Kwon et al., 2014; J. Zhou, Thompson, et al., 2013). Reported position allows one to determine the interocular contrast difference that supports equal contribution from each eye, and it is this contrast difference that quantifies “binocular balance”. Another

test embeds contrast manipulation within a global processing task (Black et al., 2011; Hamm, Chen et al., 2017; J. Li, Hess et al., 2013; J. Li et al., 2010; J. Li, Thompson, et al., 2013; Mansouri, Thompson, & Hess, 2008; Spiegel, Li, et al., 2013). In motion coherence paradigms (J. Li et al., 2010), the observers are required to discriminate the direction of a pattern of moving dots, comprised of signal-dots (moving in one direction) and noise-dots (moving in random directions). The most commonly used version of this approach for quantifying SED requires two phases: (1) determining the proportion of signal to noise dots supporting reliable direction discrimination and (2) determining the contrast of the signal-dots required to maintain performance under dichoptic conditions (Black et al., 2011). This paradigm has also been used with global orientation tasks (J. Zhou, Jia et al., 2013). In all such tests, observers perceive a coherent cyclopean percept and the relative contrast supporting optimum performance is termed the interocular 'balance point' and quantifies SED.

A second category of tests use rivalrous stimuli to quantify SED. When different monocular images fall on corresponding retinal locations of the two eyes this can lead either to an experience of *binocular rivalry* (an alternation of percept between the stimuli presented to the two eyes; Wheatstone, 1838), or to *diplopia* (double vision). The nature and extent of conflict between the information from the two eyes is a measure of the degree of SED. Bossi *et al*. (2017) quantified this using dichoptic stimuli where each eye was presented with opposite contrast-polarity versions of the same symbol (Figure 32, test #3), whereas Kwon et al presented observers with spatially overlapping rivalrous letter pairs of differing contrast (Figure 32, test #2; Kwon et al., 2015). In both cases, the objective was to measure the 'contrast balance point', or the contrast mixture required for observers' percept to be equally likely to be driven by either eye's view. This balance point (like the non-rivalrous tests) quantifies SED.

Although all tests described use centrally presented stimuli, SED has been measured at multiple locations in the visual field. Xu et al mapped SED over 17 locations (at the fovea, 2 and 4 deg. eccentricity) and report gradual variation across the field (Xu, He, & Ooi, 2011). Hess and colleagues have measured contrast balance in patients with strabismus and report that patients exhibited stronger suppression across the field than controls (Babu, Clavagnier et al., 2017). Development of this, and many of the tests described, has been driven by research exploring the role of interocular suppression in amblyopia (D. H. Baker et al., 2008; C. B. Huang et al., 2009; Mansouri et al., 2008). Several studies have examined the use of training (or *perceptual learning*) to shift balance towards the amblyopic eye as a treatment for amblyopia (E. E. Birch et al., 2015; Ooi, Su et al.,

2013). Although the benefits of such training for acuity in the amblyopic eye are clear, it is less clear if these benefits result from a shift in SED (Bossi et al., 2017; Vedamurthy, Nahum, Bavelier, et al., 2015) as has been reported (Black et al., 2012). Although the functional significance of changes in SED is debatable, it is the case that such changes can arise from short term manipulation of binocular sensory experience (e.g. through patching) in both amblyopes (J. Zhou, Thompson, et al., 2013) and controls (J. Zhou, Clavagnier, & Hess, 2013; J. Zhou, Reynaud, & Hess, 2014b).

There have been several efforts to translate lab-based measure of SED to the clinic. Typically, research groups have adapted their own method for clinical use - making tests more convenient/shorter/simpler (e.g. Black et al., 2011; Kwon et al., 2014). Here we compared a variety of dichoptic stimuli using a common psychophysical protocol. We assessed resulting SED scores based on 1) *Reliability* (consistency between two measurements made on the same participant), and 2) *Validity* (the ability of the test to capture individual differences).

We selected (or created) tests suitable for a single, rapid psychophysical protocol (effectively precluding 2-stage paradigms and tasks involving multiple stimulus locations). We ensured tasks probed a range of visual processes; from low-level (e.g. phase; Ding & Sperling, 2006), to high-level processing of spatial form (e.g. letter identity; Kwon et al., 2015), as well as motion perception. We also set out to quantify the impact of rivalry on our measures, focusing on stimuli involving no-rivalry (Ding & Sperling, 2006), contrast-polarity rivalry (Bossi et al., 2017) and spatial-form rivalry (Kwon et al., 2015).

5.2 Methods

5.2.1 Participants

We recruited thirty adult participants (18 female; 22-55yrs old) through the University of Auckland-Optometry clinic via email and poster advertisements. Based on pre-test screening, none had significant disorders of vision and all wore optical correction as necessary. Three of the observers are authors; others were not informed as to the purpose of this study. All gave informed consent prior to participation. Experimental protocols complied with the Declaration of Helsinki and were approved by the University of Auckland Human Research Ethics Committee.

5.2.2 Apparatus

Stimuli were presented on a linearised LG 21” LCD 3D monitor, with a 1920x1080 pixel resolution operating at 120Hz. Stimuli were generated in Matlab (Mathworks Ltd) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007). During testing, stimuli were viewed through wireless LCD shutter glasses (nVidia Corp., Santa Clara, CA) allowing independent control of the image presented to each eye (effective framerate was 60Hz per eye). Shutter glasses were worn over optical correction when necessary. A Minolta LS110 photometer (Konica-Minolta Ltd) was used to calibrate the monitor-luminance, using measurements made through the shutter glasses. Participants viewed the screen at a viewing distance of 100 cm to produce a pixel density of 55.9 pixel/deg.

5.2.3 Stimuli

Red-green anaglyph versions of the stimuli are depicted in Figure 32, and described in Table 7. Stimuli appeared against a mid-grey background, and were surrounded by a “vergence-lock” frame (visible to both eyes) made of black and white alternating bars; this promoted fusion.

All tests involved presenting a pair of images independently to each eye. **Test #1-2.** Stimuli comprised a pair of different letters, selected from 10 ETDRS Sloan letters (Ferris, Kassoff et al., 1982). The width and height of letters was 2.75° and the letters were the same (positive) contrast polarity. In test #1 letters were spatially separated (minimising rivalry) whereas in test #2 they were superimposed, promoting rivalry between the two different letters. **Test #3** Letter font, size and separation were identical to test #1. However, each of the two letters was comprised of superimposed pair of opposite contrast polarity letters (with the same identity) presented separately to the two eyes. In one stimulus-letter the light component went to the left eye, the dark component to the right, and in the other stimulus-letter, vice-versa. This led to strong rivalry, driven by the conflicting contrast-polarity of the components of each letter. **Tests #4-5.** Stimuli were similar to Ding & Sperling (2006) and were comprised of two horizontal sine-wave gratings with the same spatial frequency (SF; 0.52 c/deg), presented within a 2.94° square window. The two gratings were presented separately to each eye and had phases of ±45° (test #4, 90° phase difference, Figure 32-4) or ±90° (test #5, 180° phase difference, Figure 32-5), where a 0° phase grating would present a horizontal dark bar centrally positioned (on the y-axis) bisecting the square window. On a given trial the positive-phase grating was assigned randomly to one eye, the negative-phase grating to the other. In the 90° condition (test #4) addition of gratings led to a phase shifted

cyclopean percept without rivalry, whereas in the 180° condition gratings resulted in rivalry (driven by a polarity conflict similar to test #3). **Tests #6-7.** Stimuli were comprised of two 13.75° square filtered-noise patterns drifting in opposite directions (upwards or downwards). Stimuli were generated by filtering Gaussian random noise with a log Gabor filter which passed all orientations but a range of SFs (filter had a log-Gaussian profile: mean SF of 6 c/deg, bandwidth 0.5 octaves). **Test #8.** Stimuli were stereograms - stereo-defined Sloan letters (5.5° square) - defined by interocular horizontal shifts of the pixels within carrier images (435 X 435 black-white pixel arrays with each patch subtending 7.8°). Stimuli were shifted with sub-pixel accuracy using bilinear interpolation.

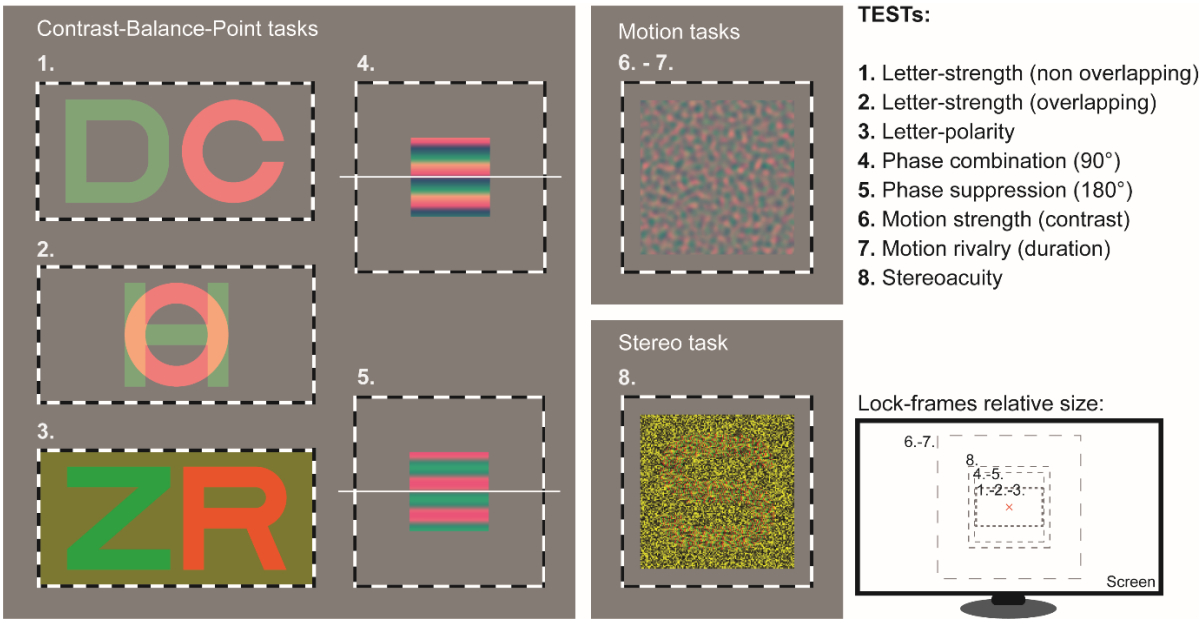


Figure 32 SED tests stimuli. (Grey panels) Examples of stimuli used in the eight tests (for the purpose of demonstration, stimuli are presented in red-green anaglyph versions). (Inset, lower right) Relative sizes of stimuli.

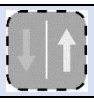
Test #	Test name	Symbol	Cue	Rivalry? (type)	Question	Measure	Primary reference
1	***Letter-strength (non-overlapping)***		*form*	×	*Which letter appears stronger?*	Contrast-BP	Kwon et al. (2015)
2	***Letter-strength (overlapping)***		*form*	✓ *(form)*	*Which letter appears stronger?*	Contrast-BP	Kwon et al. (2015)
3	***Letter-polarity***		*form*	✓ *(polarity)*	*Which letter appears whiter?*	Contrast-BP	Bossi et al. (2017)
4	***Phase combination***		*position*	×	*Does middle dark bar fall above or below horizontal ref. line?*	Contrast-BP	Ding and Sperling (2006)
5	***Phase suppression***		*position*	✓ *(polarity)*	*Does middle dark bar fall above or below horizontal ref. line?*	Contrast-BP	Ding and Sperling (2006)
6	***Motion strength (contrast)***		*motion*	✓ *(motion)*	*Does the pattern move up or down?*	Contrast-BP	Black et al. (2011)
7	***Motion rivalry (duration)***		*motion*	✓ *(motion)*	*Hold ↑ key when pattern moves ↑ & ↓ key when it moves ↓*	Time-ratio BP	Dieter, Sy, and Blake (2016)
8	***Stereoacuity***		*form*	×	*Identify the letter you see*	Threshold disparity	Frisby et al. (1996)

Table 7 Summary of eight tests. BP= Balance Point (either contrast-BP in tests #1-6 or time-ratio BP in test #7).

5.2.4 Procedure

We ran participants through eight tests, twice (to allow us to estimate test-retest reliability). Tests were always administered in the same order (#1-8). Participants first performed all tests once (run #1) and then repeated the whole test-sequence (run #2). In all tests except #7 (motion rivalry) stimuli appeared for a maximum of 4.5 s, after which the display switched to showing the response-choices for a maximum of 10s. Observers could respond before the display of response options, which triggered presentation of the next trial. Participants signalled their response verbally, which the experimenter recorded using the computer keyboard. Note that no feedback was provided in any test (since we are estimating bias in all but test #8). Each test took approximately 90s to administer. The experimenter instructed the participant before each test as highlighted in Table 7. Each test took no more than 1 min for participants to perform and was separated from the upcoming test by a 1-2 min break. During this time the experimenter described the next test and participants were free to look around the room. This schedule minimised any impact of testing on subsequent tests (e.g. because of adaptation). Total duration of the experiment - incorporating 8 tests x 2 repeats, instructions and breaks - was around 50 min. Tests #1-7 measured a Balance Point (BP) from 0 (participant used

their left eye only), to 1 (right eye only). For tests #1-6 BP was based on contrast, and for task #7 it was based on a ratio of time spent experiencing the stimuli presented to each eye.

Tests #1-6 quantified the ratio between the contrast-level applied to each dichoptic-image that leads to either of the two images (called A, B below) being equally likely to be chosen. This contrast-BP was determined in 20 trials using an adaptive staircase algorithm (QUEST; A.B. Watson & D.G. Pelli, 1983) which, for a stimulus comprised of the mixture of images *A* and *B*, converged on a threshold (α) producing 50% identification of (tests #1,2,4-6) image *A* or (test #3) the "whiter" stimulus. The values QUEST produced were clamped in the range 0.0-1.0 (where 0.0 is exclusive presentation of the cue to the left eye, 0.5 equal presentation to left and right eye and 1.0 exclusive presentation to the right eye). QUEST's guess-rate (γ) was set to 0.5 (2 AFC), the lapse rate (λ) to 0.01 and the slope-estimate (β) to 3.5. The initial guess for threshold or contrast-BP was 0.5 (i.e. balance) with an associated standard deviation of 0.7. For tests #1,2,4,5 and 6, QUEST set the contrast (*C*) of the right eye component (C_{right}) and C_{left} was set to 1-C_{right}, for test #3 contrast manipulation is described in the next section. Note that QUEST values were determined from a Monte Carlo simulation on a population of ideal observers whose simulated-balances spanned the range 0.05-0.95 and whose other psychometric characteristics (β, λ) were taken from Kwon et al. (2015) – more information in section 5.2.5. **Test #7** consisted of 1 minute of exposure to rivalrous motion. Either upward or downward motion was randomly assigned to the left and right eyes for the initial 30s test period and then direction was switched across eyes for the remaining 30s (to counter-balance any bias for a stimulus in a given direction rather than from a given eye). During stimulus presentation, observers pressed and held either the U ("up") or D ("down") button on the computer keyboard to indicate their dominant percept. In case of uncertainty, e.g. due to a "patchy" percept, the participant was instructed to report the more dominant direction. **Test #8** estimated a stereo-threshold (not bias) for letter-identification, using a 20-trial QUEST staircase. QUEST's guess-rate (γ) was set to 0.1 (10 AFC), the lapse rate (λ) to 0.01 and the slope (β) to 3.5. The initial guess for threshold was set to 3'20".

Details of contrast manipulation for test #3 For test #3, the QUEST value set the luminance of the components of the opposite contrast polarity left/right-eye letters forming each of the two stimulus letters. First, the highest increment was randomly assigned to the stimulus-letter A or B. Then, the luminance of the light component-letter in one stimulus-letter and the dark component-letter in the other stimulus-letter were set

to be equal increments and decrements (range ±50cd/sq.m; example in Figure 19; taken from Bossi et al., 2017). Thus, for a background luminance of 50 cd/m^2 if QUEST produced a value of 0.8 then the light component letter in, for example, stimulus letter A was 50+0.8*50=90 cd/m^2 and so the dark component letter in stimulus letter B was 50-0.8*50=10 cd/m^2. The dark component letter in stimulus letter A was 50-(1.0-0.8)*50=40 cd/m^2 and the light component letter in the stimulus letter B was 50+(1.0-0.8)*50=60 cd/m^2. In the extreme case of QUEST producing the limit value of 0.0, then the light component letter in, for example, stimulus letter A was 50+0.0*50=50 cd/m^2 and so the dark component letter in stimulus letter B was also 50-0.0*50=50 cd/m^2. The dark component letter in stimulus letter A was 50-1.0*50=0 cd/m^2 and the light component letter in the stimulus letter B was 50+1.0 *50=100cd/m^2. Conversely, when QUEST converged to 1.0, the dark component of A and the light component of B were matched (at 50 cd/m^2) while A had a light component at 100 cd/m^2 and B a dark component at 0 cd/m^2.

5.2.5 Response Procedure: simulation

The test method used was first optimised using Monte Carlo simulation of a population of ideal observers with uniform distribution of average true BP from 0.0 to 1.0 (as indicated in legend Figure 33d). This is unlikely to represent the distribution of contrast-balance in the non-clinical population but we would like our test to operate effectively regardless of the true BP.

Response characteristics of ideal observers (e.g. slope of psychometric function) were derived from Kwon et al. (2015).

We ran 1024 simulations using a bootstrap method (around the ‘True balance’ fitted values, listed in legend in Figure 33d).

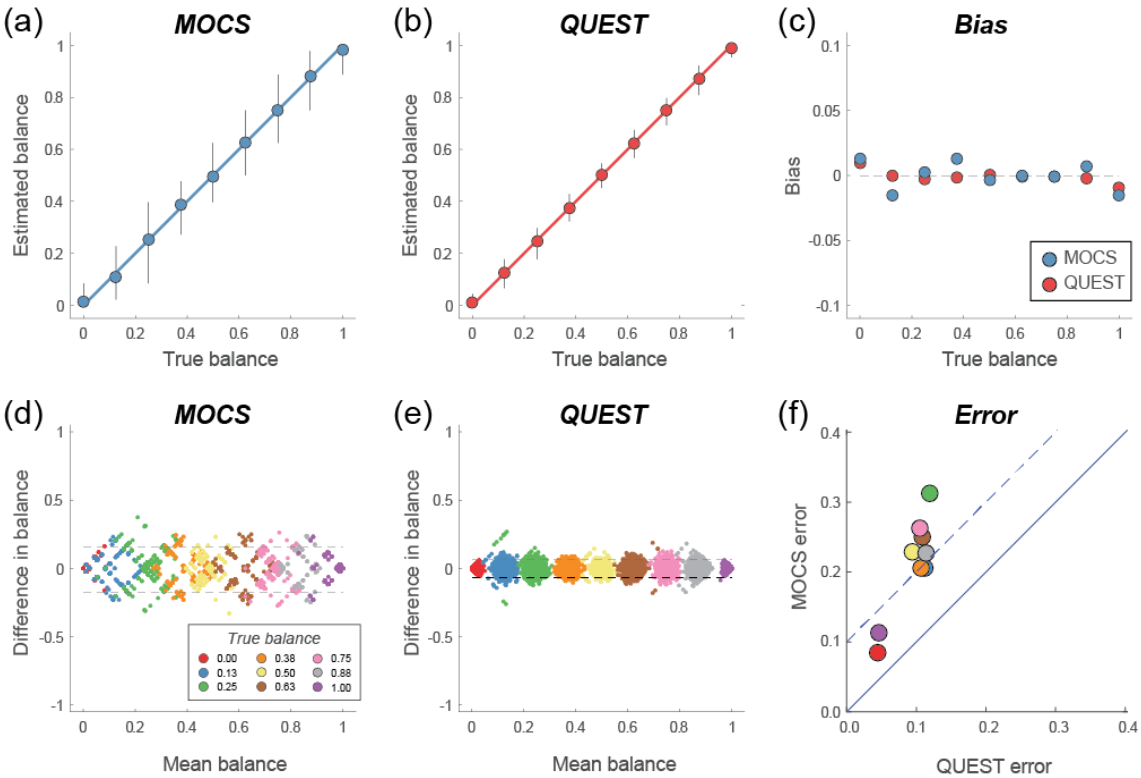


Figure 33 Comparison of simulated estimates of contrast-balance derived from repeated testing of a set of ideal observers over 20 trials. Simulated data were measured using either (a,d) Method of Constant Stimuli or (b,e) QUEST staircase procedures. (a,b,c) Both methods yield largely accurate estimates although note QUEST-estimates have smaller error bars on (compare a to b) and are (c) less biased than MOCS-estimates. (d,e) Bland-Altman plots of simulated test-retest reliability confirm superiority of the QUEST procedure. (f) 95% confidence interval on mean QUEST estimates of contrast-balance (averaged values, around the true estimates, excluding the extremes of the values range) are around 0.1. Thus, a measured difference exceeding 0.1 is statistically significant.

5.2.6 Analysis

For tests #1-6, BP was derived by fitting the binary response data (using each eye at a time) from the 20 trials of a single QUEST-controlled run with a cumulative normal function using the Palamedes toolbox (Prins & Kingdom, 2009). Further fitting was performed in order to quantify confidence in the BP estimate. In order to do this, we pooled stimulus-levels (and their associated responses) into four bins, assumed binomially distributed errors at each level and bootstrapped fits to these binned data. Although this is not an exact estimate of error (since we cannot bootstrap the continuous data used to derive the contrast-BP) it is nonetheless a useful indicator of confidence in these estimates.

For test #7 we quantified BP as the proportion of left-eye dominance, using the proportion of frames (within a given trial) when the participants' response was consistent with their relying on the left-eye-view during exposure to rivalrous stimuli. We excluded responses occurring within 4.2 s (i.e. 500 frames) of the beginning, and the middle (when

directions switched between eyes) of the 60s sequence. Responses composed of simultaneous keypresses were excluded from analysis.

We first describe BP for each individual and each test by averaging run 1 and 2 together, and also calculate the mean BP for each participant across all tests. To estimate test-reliability we calculated Bland Altman 95% limits of agreement, or Coefficient of Repeatability (CoR - lower values indicate higher repeatability; Vaz, Falkmer et al., 2013) as well as the Intraclass Correlation Coefficient (ICC – higher values indicate higher repeatability), based on a mean rating (k=2), absolute agreement, 1-way random-effects model (McGraw & Wong, 1996). In other words, the CoR measures the consistency of observations within individuals, while the ICC quantifies the extent to which repeated measurements from the same individual are in agreement compared with the variation between individuals. The best possible ICC is =1, indicating a complete reliability of a test. On the other extreme, ICC=0 indicates that there is no more agreement between repeats from the same individual compared to those from different individuals, and therefore the test used has a very low reliability. We examine how CoR and ICCs are influenced by the number of trials run (by re-calculating BP with Palamedes refits for fewer trials), and investigate the reliability of tests by comparing measures to one another. We note that highly reliable outcomes can arise from a measure which has low test-validity, being insensitive to differences in SED between individual (in other words, a test which always elicited the same outcome would be highly reliable, but not very useful). The absence of a gold standard measure of SED means we cannot directly assess which test is the most accurate by comparing it to such a standard. We therefore relied on measures of test-retest ranks, i.e. Mean Average Precision (MAP) and Fractional Rank Precision (FRP), to incorporate both reliability *and* validity (Dorr, Elze et al., 2017). MAP indicates how likely is a test to score a result for a certain individual as truly distributed in the general population (e.g. how likely is the person with the second-best SED to score the second closest score to 0.5?). FRP, instead, describes the reliability of a test score to be replicable considered the variability of the tested population, from 0.5 (chance) to 1 (perfect test-retest). Finally, we use regression to compare SED data to stereoacuity. FRP employs an information retrieval approach and evaluates a test by quantifying how identifiable a participant is from their set of test-scores.

5.3 Results

Tests #1-7 estimated observers Balance-Point (BP): #1-6 estimated observers' contrast Balance-Point, #7 quantified rivalry; test #8 measured stereoacuity. We assessed tests according to their 1) *reliability* (consistency between two measurements of the same test made on the same participant), and 2) *validity* (the ability of the test to capture individual differences amongst participants). Reliability was quantified using the Bland Altman Coefficient of Repeatability (CoR; the variance of the difference between runs 1 and run 2). The other metrics in Table 8 provide statistical estimates of both reliability and validity.

Are observers more reliant on their left or right eye? Figure 34a plots the mean BPs (of the two runs) of tests #1-7 for all observers. Participants IDs are ordered along the abscissa by BP (averaged across the two runs). The overall mean-BP (for all participants; Green symbols), obtained by averaging mean contrast-BP, for tests #1-6 *and* mean proportion of instances when the percept was determined by the left eye (test #7) was 0.52 (σ=0.04). Mean contrast-BP (excluding test #7) was 0.52 (σ=0.05). Note the clustering of BP results around 0.5, indicating near-perfect binocular balance; as one would expect for observers with normal binocular vision. In line with previous studies, we report a prevalence of right eye-dominance of 21/30 (mean BP=0.54, σ=0.03). 9/30 participants were mainly left-eye dominant (mean BP=0.47; σ=0.02). Sixteen participants showed a marked eye-preference in all balance tests (only five had a left-eye preference). Participant were classified as having a marked eye-preference if they showed both:

1) A consistent eye-preference across tests. Specifically: $(|\mathbf{meanBP\ across\ tests}|) > \left(\mathbf{0.5} \pm \frac{\sigma}{2}\right)$
2) No one test could indicate a clear preference for the eye opposite to that indicated by (1). i.e. exclusion if BP on a given task >0.5+2σ (if 1 indicates left eye dominance), or <0.5-2σ (if 1 indicates right eye dominance).

Criteria 1) and 2) used the standard deviation (σ =0.021) of the mean test-retest difference between runs, averaged between participants and across tests #1-7.

Tests #1, 3, 4, 5 and 7 elicited BPs more similar to one another (and to the overall mean balance) compared to tests #2 and 6 (Figure 34a). Figure 34b plots a sample of data from three participants performing two runs of the seven balance tests. The dashed line denotes perfect binocular balance (BP=0.5). The green line indicates the mean estimated balance-level across the seven tests (with exact values provided in the green

box above each plot). We found a generally high level of agreement across runs of the same test: the absolute difference was 0.05±0.025 between participants (averaging, for each participant, the difference across tests) and 0.05±0.028 between tests (averaging, for each test, the difference across participants). We examined the consistency of balance estimates using a correlation analysis. Specifically, for each task we correlated the series of contrast-BPs (across all observers) for one task (averaged across runs 1 and 2) with the series of contrast-BPs (across all observers) averaged across both runs and all other tasks. We observe consistency of contrast-balance estimates made using different tests as indicated by the mean correlation coefficient (mean R= 0.564, σ=0.102, *p*=0.0012) obtained by averaging the R from each task.

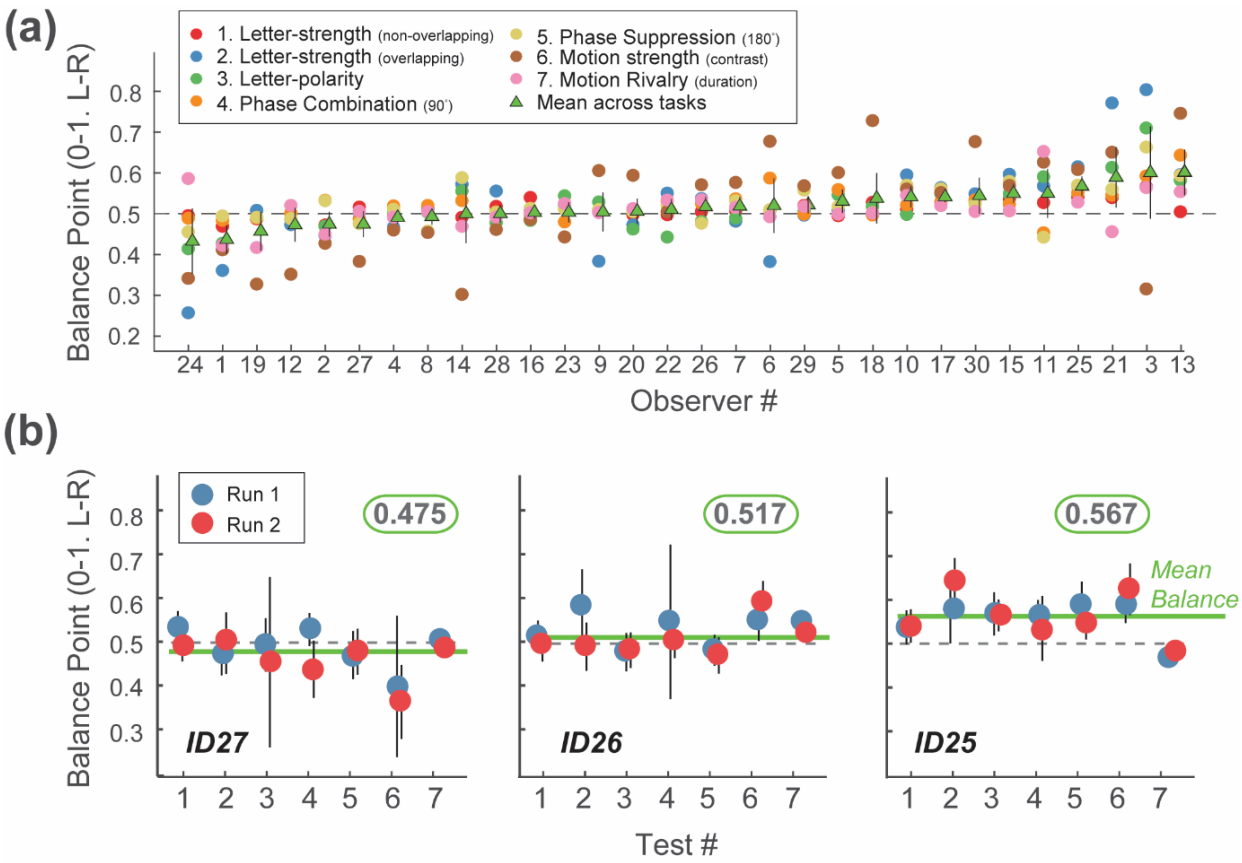


Figure 34. (a) Estimated balance point (BP) for adult participants with normal vision performing tests #1-7, ordered by mean BP across tests (green triangles; error bars are 95% confidence intervals). Here, BP=0.0 would indicate complete reliance on the left eye, 1.0 complete reliance on the right eye and 0.5 equal reliance on each eye (binocular balance). Tests are colour coded according to the legend. **(b)** Illustrative results from three participants, chose from the left, central and right batches of IDs on the abscissa in (a). Blue and red symbols show data from the first and second run respectively, and error bars are 95% confidence intervals on estimates. Mean BP across tasks (green line) is the green-boxed figure in each panel.

Which is the "best" test? We assessed 'best' in terms of (a) how reliable the tests are (across two runs) and (b) how well they capture individual variation in SED across our group. In other words, the best test minimizes the range of measures across runs (narrow range on the y-axis), and assuming that SED varies within the population tested, maximizes the range of measures across observers (wide range on the x-axis). To visualise this, Figure 35 shows Bland Altman plots of data from the seven balance tests as well as test #8: stereoacuity (n.b. data are plot on different axes to other sub-plots). It is clear that some tests such as #6: motion strength, lead to a wide range of balance estimates across observers but also to high variability of balance-estimates across runs (see also Figure 34). Conversely, tests such as letter-strength (Figure 35, task #1) elicit both a narrower range of balance estimates and much lower variability across runs. However, the high degree of repeatability of this test arises from it yielding a BP estimate of 0.5 for almost all participants, suggesting it is unable to differentiate subtle difference in SED (i.e. it has poor test-validity). We note that it is the closest test we have to the *Sbisa bar*, in that there is not spatially overlapping information and participants are required to make a judgment of contrast.

A variety of statistics quantify this and a selection are given in Table 8. Test #3 (letter-polarity) maximises intra class correlation (ICC), *F*, mean average precision (MAP), and fractional rank precision (FRP). The Coefficient of Reliability (CoR) is lowest for test #1 - the letter-strength (non-overlapping) task. However, as indicated above, it would appear that this reliability comes at the expense of failing to capture individual variation in SED. Based on ICCs and corresponding F values, tests #2 and #3 appear particularly useful. However, ICCs are driven by values at the ends of the measured range, making them susceptible to outliers. In our data set, participant ID3 reported quite unbalanced but reliable scores on tests #2, 3 and 5. This individual may be an outlier, inflating ICC scores for these tests. MAP and the related FRP estimates circumvent this issue by scoring on rank rather than absolute value. This difference in method impacts test #6 the most, as it has a poor ICC (and *F*), but fair MAP and FRP values.

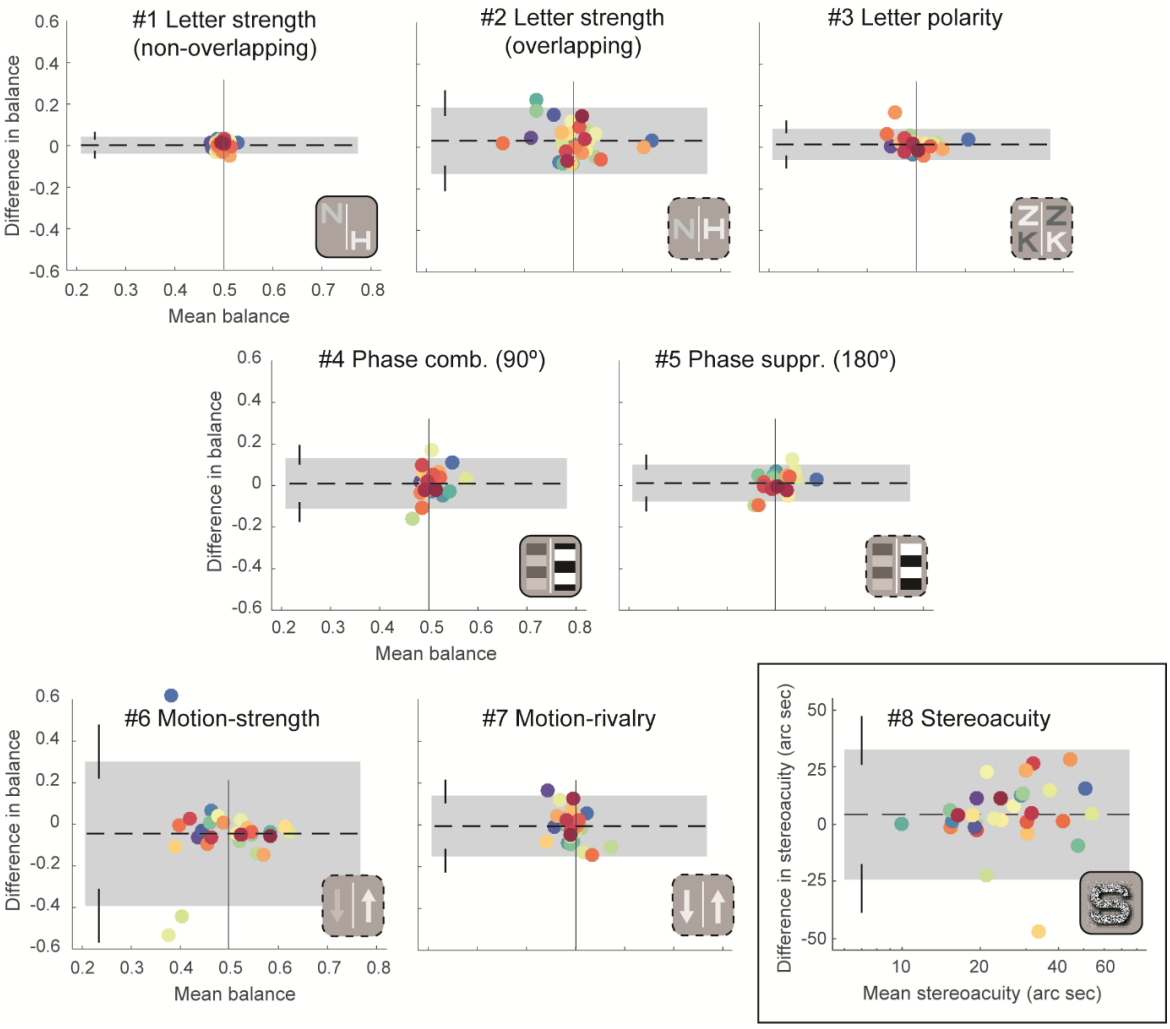


Figure 35. Bland-Altman plots of results from the eight tests. (#1-6) Mean contrast balance across runs is plotted against the difference in contrast balance. (#7) Relative time spent in each possible rivalrous state plot against the different in this measure across runs. (#8) Mean stereoacuity plot against the difference in stereoacuity across runs. Shaded regions denote the 95% confidence intervals across estimates with error bars on these estimates calculated using the procedure given by Carkeet (2015). Note high repeatability in (#1) but restricted range of estimated balance across participants. Motion test (#6), in contrast, produced a broader range of estimated balance but at the expense of repeatability. (#3) The letter-polarity test combines high repeatability with a broad range of estimated contrast-balance.

	ICC	*F (p)*	*CoR*	*MAP*	*FRP*
1. Letter-strength (non-overlappin	0.603	4.035 (0.00014)	**0.058***	0.174	0.638
2. Letter-strength (overlapping)	0.727	6.336 (<0.00001)	0.224	0.197	0.648
3. Letter-polarity (overlapping)	**0.803***	9.152* (<0.00001)	0.105	**0.299***	**0.776***
4. Phase-comb (90°)	0.230	1.597 (0.10425)	0.169	0.170	0.592
5. Phase-suppression (180°)	0.629	4.388 (0.00006)	0.124	0.198	0.711
6. Motion strength (contrast)	0.316	1.924 (0.03988)	0.494	0.261	0.752
7. Motion rivalry (duration)	0.201	1.504 (0.13607)	0.208	0.141	0.539
8. Stereo	0.392	2.29 (0.013)	40.2	0.222	0.688

Table 8. Adults with normal vision: Summary statistics for performance of the eight tests. *Indicates best performance (across tests #1-7) for a given metric. *ICC*:intraclass correlation, *F*: ANOVA *F* statistic, *CoR*: coefficient of repeatability (lower is better), *MAP*: mean average precision (higher is better), *FRP*: fractional rank precision (higher is better; Dorr et al., 2017). Note CoR measures for Stereo are unique in being expressed in arc sec.

Figure 36 shows the variation of the CoR and ICC over run-length. Here we analysed from 25 to 100% of data (i.e. either trials 5-20, for tests #1-6 and 8 or from 4.2-30s and 34.2-60s in test #7), to limit the confounding role on results of a possibly uncertain performance at the beginning of each test (e.g. due to an initial lack of familiarity with the task). For each test (colour coding given in the legend), CoR-coefficient of repeatability (Figure 36a) and ICC-interclass correlation (Figure 36b) indexes are obtained by comparing first and second run across all observers. Predictably, there is better agreement towards the end of runs, as evidenced by the descending (Figure 36a) or ascending (Figure 36b) trend of the plotted lines. Figure 36c visually represents CoR (x-axis; lower is better) against ICC (y-axis; higher is better) indexes for data from the whole run for all of the tasks tested. Note the clustering of results. Results from the two motion- and the non-rivalrous phase-combination tests were substantially poorer than the other tests.

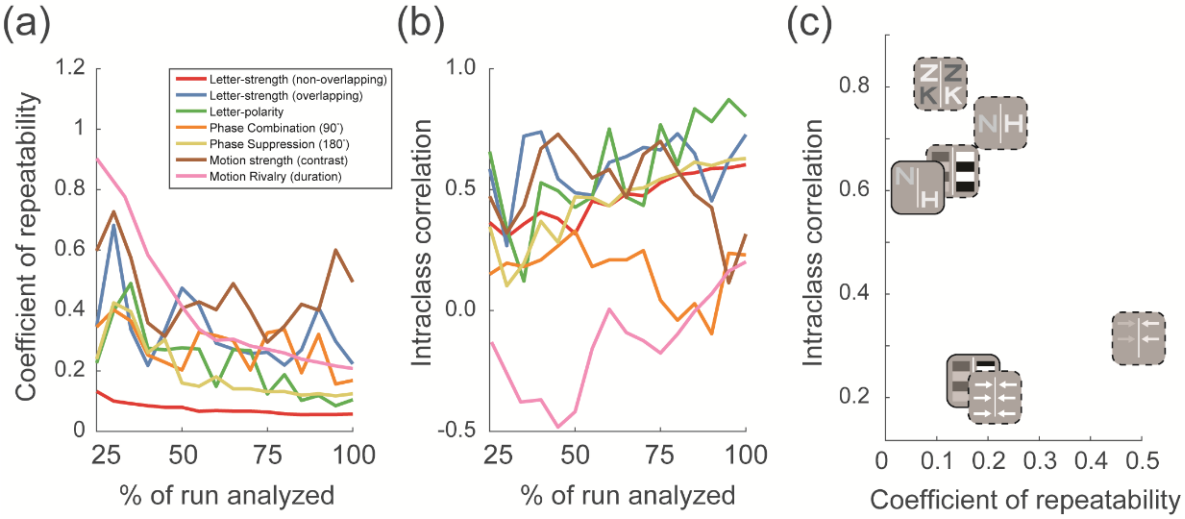


Figure 36. CoR and ICC in adults with normal vision. Plot of the evolution of (a) coefficient of repeatability (lower is better) and (b) interclass correlation (higher is better) over run-length (expressed as a percentage of the run analysed, e.g. 100%=20 trials) for the seven balance tests. The tests are listed in legend-panel (a). (c) Plot of intra-class correlation against coefficient of repeatability.

We next assessed whether increased binocular imbalance was associated with poorer stereoacuity and/or with more marked eye-dominance in rivalry. As shown in Figure 37a, we fitted a regression line to the data points corresponding to (y-axis) the magnitude of SED and (x-axis) the stereoacuity estimate from the same participant (mean of the two runs). The mean magnitude of SED was obtained by averaging contrast-BP indexes across tests #1-6 and computing the absolute difference of this value from 0.5. The fit was consistent with only a marginal association between measures (R=0.30), resulting not significant (p= 0.104). Figure 37b plots SED magnitude against the mean duration of instances of perceptual dominance, as measured in the rivalry test (#7). These data show a non-significant correlation of R=0.19, p=0.302. Participants' rivalrous percepts lasted for an average of 2.27seconds (σ=1.42s; median=1.73s).

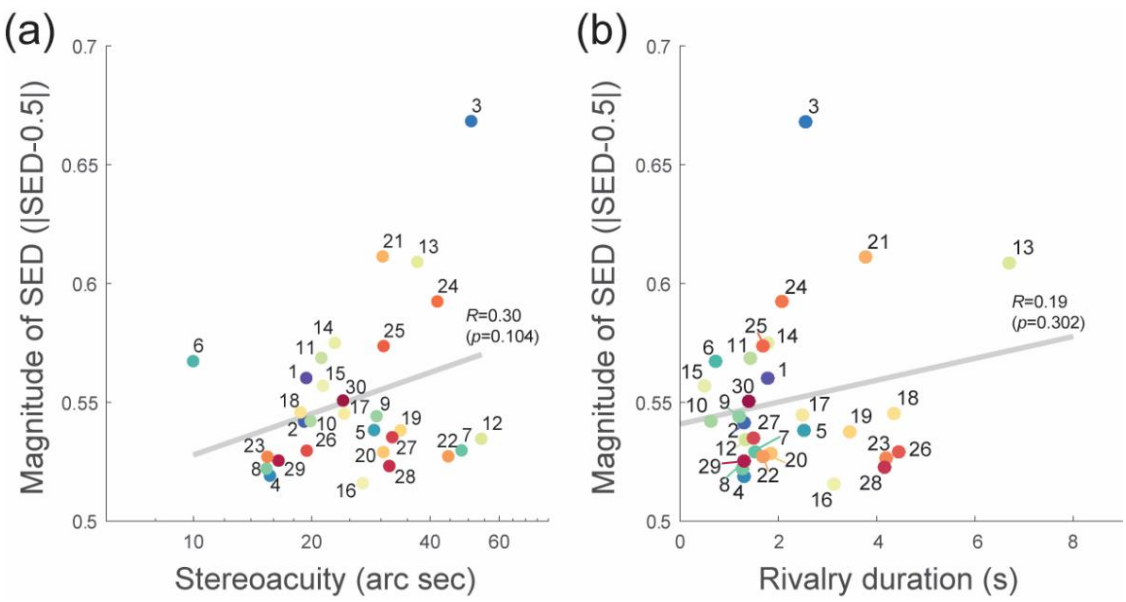


Figure 37 SED and stereoacuity. Data from each numbered participant (colour coding of tests is as Figure 3) showing either **(a)** stereoacuity (task #8) or **(b)** the duration of perceptual dominance (task #7) plot against magnitude of SED (based on the mean contrast-Balance-Point averaged across test #1-6). All data are averaged across two runs. Magnitude of SED is |SED-0.5|, i.e. 0.5=balanced vision, and 1=complete dominance of one eye. On each graph, a regression line has been fit to data, and correlation indexes (*R*) with associated *p*-values are reported.

5.4 Discussion

We compared eight tests that use dichoptic stimuli to quantify binocular visual function. We measured sensory-eye-dominance (SED) in tests #1-7, and stereoacuity in test #8 in 30 individuals with ostensibly normal vision. All SED tests involved estimation of a *Balance-Point* (BP): an estimate of the extent to which participants relied equally on the two components of a dichoptic stimulus-pair. Specifically, pairs of *contrast*-modulated stimuli were used in tests #1-6, comprising Sloan letters (adapted from Bossi et al., 2017 -test #3; Kwon et al., 2015 -tests #1-2), reciprocally-shifted gratings (from Ding & Sperling, 2006 -tests #4-5) or up/down drifting noise-patterns (loosely based on the motion task by Black et al., 2011 -test #6). Test #7 (inspired by Dieter et al., 2016) measured rivalry, as the proportion of dominant-eye instances (presenting the same pairs of stimuli used in # 6 but at a fixed contrast). In addition, test #8 measured stereoacuity, using stereo-defined Sloan letters (variant of Frisby et al., 1996).

We observed a high level of consistency in BPs across tests for each participant. The mean magnitude of SED, obtained by averaging absolute BPs, was 0.55, σ=±0.03, with the majority of participants (22/30; mean BP=0.5; σ=0.03) showing right-eye sensory dominance, in line with previous results (e.g. Ehrenstein, Arnold-Schulz-Gahmen, &

Jaschinski, 2005; Johansson et al., 2015; Jonathan S. Pointer, 2012). Overall, regardless of the statistics used, test #3 showed the best performance in term of validity and reliability; while tests #7 and #4 were the poorest (see Table 8). We found no significant correlation between stereoacuity and SED measures: better binocular balance (i.e. BPs around 0.5) was not associated with better stereoacuity Figure 37a. Such a null result could simply arise from a lack of variation in stereoacuity amongst our (normally sighted) observers and indeed it has recently been observed that low between-subject variability elicited by robust psychophysical tasks such as stereoacuity, may make them ill-suited for studying individual differences (Hedge, Powell, & Sumner, 2017).

Why are some tests better than others? In terms of test-validity an ideal test captures a range of BP estimates across participants. We note that binocular imbalance manifests most robustly for stimuli presented at the same location so that estimates of BP from test #1 (non-overlapping letter strength) converge on 0.5. This fundamentally limits the validity of this test for capturing variation in SED within a population of people with normal vision.

In terms of test-reliability, one might expect that tests involving rivalrous stimuli would elicit less reliable responses from participants as e.g. rivalrous percepts typically fluctuate in time. We presented stimuli for relatively short durations, to limit such perceptual alternation. Nevertheless, our expectation - that these stimuli would still elicit more random response – was not borne out by the reliability of tests involving rivalrous stimuli. Rather, our finding of high levels of reliability on these tests suggests that the SED may selectively drive the early stages of a rivalrous percept. This is consistent with earlier reports of participants' first percept of a rivalrous display predicting their eye dominance (e.g. quantified using percept frequency) calculated for stimuli that alternated over dozens of seconds (Dieter et al., 2016; Xu et al., 2011).

Among the contrast-balance tests, those which rival in motion (#6) and form (#2) show the poorest test-retest reliability (CoR). We believe that the different performance on tests #3 and 5 compared to tests#2 and 6 arises from polarity-rivalry removing the need for participants to explicitly judge relative stimulus strength. For two rivalrous, same-polarity but different-identity targets (e.g. test #2), observers likely judge relative stimulus-strength by comparing two independent estimates of contrast. In this case *binocular imbalance manifests as a difference in magnitude of the two cues* (i.e. one will be "stronger/whiter/brighter" than the other). In contrast, the appearance of two rivalrous,

opposite-polarity but same-identity targets (test #3), alternates over time with the *target-pair* appearing as either left-black/right-white or left-white/right-black. In this scenario *binocular imbalance manifests as a difference in probability that one state will dominate over the other.* In other words, for test #2 the observer is faced with a judgement of a subtle difference in perceived luminance, whereas in test #3 they are always judging which of two (apparently) black and white letters is white. This task is straightforward for our adult participants and sufficiently simple for children, in fact we could reliably administer a very similar test using an unsupervised measurement system, as part of the home-based binocular amblyopia therapy, described in Chapter 3 – and published in Bossi et al. (2017).

How could new tests fit into clinical practice? There is an unmet need for accessible, valid and reliable methods to measure binocularity in standard care. In the clinic, binocularity is frequently assessed measuring stereoacuity. However, 1-14% of the population is stereo-blind (Bosten et al., 2015), and binocular function is necessary but not sufficient for stereoacuity. Moreover, standard stereo-tests (e.g. Frisby and Randot) result in non-measurable stereoacuity in cases where compensation for visual deficits (e.g. balancing stimulus visibility across the eyes) allow patients to see stereo (Tytla, Lewis et al., 1993), or achieve binocular summation (D. H. Baker, Meese, Mansouri, et al., 2007). We therefore would anticipate SED being able to (a) quantify subtle binocular deficits in people with normal stereoacuity and (b) reveal the presence of binocular capacity in the absence of measurable stereoacuity. Our current results speak to (a) – we observe a range of SEDs in our test group – albeit centred on 0.5 or balanced vision. Differences between observers are stable, both in terms of test-retest reliability and in terms of agreement amongst different tests of SED.

Other clinical tests are problematic for their own reasons. For example, tests of “sighting-dominance” (e.g. Miles or Porta tests) deliver binary “left” or “right” measures of eye-dominance. Further, sighting-dominance can be inconsistent, influenced by gaze direction (Khan & Crawford, 2001), and by the test used (Rice et al., 2008). The only continuous estimate of eye-dominance available to clinicians comes from Sbisa bars which rely on the clinician’s judgement of what stimulus level induces a patient to report diplopia. Our procedures have much in common with the Sbisa bar but automate stimulus selection, have the patient perform a forced-choice task, and use all of the response-information to calculate the balance point. This, we believe, will contribute to superior test reliability and better compliance from patients.

It is important to note that the sight-dominant eye does not necessarily support better visual acuity (Pointer 2007), and laterality - measured with either SED or sight-dominance - only agrees in 50% of cases (Pointer 2012). Further, tests of SED (Suttle, Alexander et al. 2009, Pointer 2012) as well as those of other forms of binocularity, such as sight dominance (Rice, Leske et al. 2008) and stereoacuity (Heron and Lages 2012), produce results that differ depending on the particular visual task and test-conditions. Clearly, both sensory and motor aspects are involved in binocularity. Thus, a complete assessment of functional binocular vision requires more than one type of binocular measurement. The lack of correlation between our (reliable) measures of SED and stereo-acuity suggests these tests are complementary. Future research could more directly focus on determining the sub-processes that support complete binocular vision, and developing a more complete battery of tests to probe them.

5.4.1 Conclusion

We compared several tests for rapidly quantifying sensory-eye dominance and a test of stereoacuity. A judgement of which of two dichoptically-superimposed opposite contrast-polarity patterns dominated the participant's percept (test #3) supports a reliable and sensitive measure of sensory-eye dominance in only 20 trials. Practical and reliable measure of SED have application in amblyopia research and as part of a more thorough assessment of binocularity in the clinic.

This chapter will be extended and is due for publication (in 2019).

6 Quantitative tests of binocular vision in amblyopia

In the previous chapters, we introduced amblyopia, the complex pattern of associated deficits and the need for improving the management of this syndrome. its binocular nature has been discussed, as well as the difficulties to measure binocularity. We compared a series of different behavioural tests of sensory eye dominance & stereoacuity on a group of adults with ostensibly normal vision (see section 5.3). In this chapter, we evaluate the same tests on children and adults with amblyopia, in order to identify tests that would best complement clinical practice and compensate for the current lack of reliable and valid tests for quantifying binocularity. We also compare results with those obtained on a control group of children, in order to make a clearer evaluation of the influence of age and/or amblyopia on these measures.

6.1 Introduction

Although the mechanism of amblyopia continues to be a subject of debate, there is considerable evidence for the involvement of interocular-suppression (E. E. Birch, 2013). Suppression is associated with a **disruption of normal binocular input** (due to a degree of monocular deprivation) that induces changes in cortical-ocular dominance (Chadnova, Reynaud et al., 2017). In amblyopia, imbalanced sensory experience (e.g. due to inter-ocular difference in refractive errors and/or eye-misalignment) leads to an abnormal strengthening of excitatory- and inhibitory-neural connections, driven by the response of the healthy- and the deprived-eye respectively (Hubel & Wiesel, 1965a; Kiorpes et al., 1998; Movshon et al., 1987). Thus, the input from the amblyopic eye can be said to be suppressed. Suppression is thought to play an important role in reducing the experience of *diplopia* - or double-vision (J. Li et al., 2011). This phenomenon is important for the management of amblyopia as it is a **predictor** of the risk of developing double vision (Newsham & O'Connor, 2016). In particular, if treatment is started in adulthood when less cortical plasticity is present (Takesian & Hensch, 2013) and the ability of a patient to recover from adverse effects of treatment (such as diplopia) may be compromised (V. K. Tailor, Schwarzkopf, & Dahlmann-Noor, 2016).

It is currently unclear the extent to which ‘binocular-balancing’ treatments rely on breaking suppression (Ding & Levi, 2014; R. F. Hess & Thompson, 2015) or alternatively if reduced suppression is an incidental effect of the treatment (S. L. Li, Reynaud et al., 2015; Vedamurthy, Nahum, Huang, et al., 2015). Either way, it is important to be able to quantify the inter-ocular difference in response to sensory stimuli, in order to identify the best stimuli to train vision by rebalancing image-visibility. In other words, measuring sensory-eye-dominance (SED), e.g. by finding the Balance-Point (BP) in the signal yield by dichoptic stimuli that allows ‘normal’ binocularity, is a critical tool for the development of new treatments for amblyopia. However, as discussed in the previous chapter*, current clinical practice does not provide an objective and reliable way of quantifying sensory-eye dominance.*

Patients with a disorder of binocular vision, such as amblyopia, would be expected to show a greater imbalance in eye dominance in proportion to the severity of their condition. However, as demonstrated (Jonathan S Pointer, 2007) the eye with better visual acuity is not necessarily the sight-dominant eye, although this discrepancy likely depends on the extent of the inter-ocular difference in visual acuity (D. Zhou, Ni et al., 2017). With that proviso, in this chapter we refer to *fellow-eye* (FE) or *amblyopic-eye* (AE)/weak-eye based on visual acuity, and to *dominant-eye* (DE) and *non-dominant-eye* (NDE), based on SED.

A common underlying principle of recently emerging binocular therapies is to **re-balance the visibility** of the stimuli across the eyes (Ding & Levi, 2014; R. F. Hess & Thompson, 2015), with the aim of reducing the imbalance in SED. Most therapies use progressive re-balancing of inter-ocular *contrast* to achieve balanced-binocularity (see also section 2.8.3). However, contrast sensitivity (especially at high spatial frequencies) is significantly affected by reducing *luminance* (Campbell & Robson, 1968). For instance, by manipulating unilateral mean luminance it is possible to modulate contrast sensitivity in amblyopic viewers (D. H. Baker et al., 2008) and to reduce the imbalance in SED (Ding & Levi, 2014) while improving binocular combination (J. Zhou, Jia, et al., 2013). It has been proposed that severe and mild amblyopia could be based on sensitivity to *contrast*-modulated or *luminance*-modulated stimuli, respectively, as processes exhibit differential sensitivity to the depth of deficit (Chima, Formankiewicz, & Waugh, 2016).

With that in mind, *why do we need better tests of SED for measuring vision in amblyopia?* As discussed in the previous chapter, there is a need to expand testing of binocular

vision beyond stereopsis. In particular, in amblyopia a degree of binocularity can be present even when stereopsis is unmeasurable (D. H. Baker, Meese, Mansouri, et al., 2007; Mansouri et al., 2008; Tytla et al., 1993). A thorough evaluation should explore the many sub-processes that contribute to binocular vision: binocular combination, binocular rivalry and, finally, stereopsis (see sections 1.3, for more details on these processes, and 2.3.6, for the related deficit in amblyopia). This is particularly important, in terms of understanding why a patient with amblyopia does not have normal binocular vision, since we note that there is a huge range of individual differences in stereopsis (Robert F. Hess, To et al., 2015), binocular rivalry (Dieter et al., 2016) and stimulus rivalry[7] *in the normal population*. A wider range of tests would allow us to more accurately ascertain the source of a patients' problem. During normal visual development, stereopsis is the last visual ability to fully develop, at around the age of 5 years (Braddick & Atkinson, 2011). In fact, stereoacuity is absent if e.g. sensory fusion is unmeasurable, whereas unmeasurable stereoacuity does not necessarily indicate a lack of binocularity. It would be useful to have an efficient, practical, precise and age-appropriate set of measures to facilitate early detection of problems with binocular vision for use in the clinic and in wider (e.g. pre-school) screening programmes.

6.2 Methods

6.2.1 Participants

Inclusion/exclusion criteria

Twenty children - aged 5 to 11 yrs (mean age: 7.8 yrs, σ=1.6 yrs; 10 with amblyopia and 10 visually normal) and ten adults – aged 20 to 61 yrs (mean age: 41.3 yrs, σ=14.5 yrs; with amblyopia) were recruited at Moorfields Eye Hospital. Participants were invited to take part while waiting for a routine visual assessment at the hospital. Clinical notes (not more than 6 months old, for adults, and 3 months old, for children) were screened in advance to ensure visual status fell within our inclusion criteria. In particular, we looked for not-variable ocular motility and a conclusive result from a cover test. During this first meeting, participants and caregivers were briefed on the nature of the experiment, and consent/assent was obtained. Participants were tested immediately following their scheduled appointment. This study was approved by University College London and the local NHS Research Ethics Committee. All procedures followed the tenets of the Declaration of Helsinki.

[7] This differs from conventional rivalry in that it is triggered by dissimilar stimuli being rapidly swapped between the eyes (Patel, Stuit, & Blake, 2015).

The inclusion/exclusion criteria for participants with amblyopia (n=10 children, 3 to 12 yrs) and 10 adults (≥18 yrs), were similar to those previously described for our BBV-treatment-study (see Chapter3), with the difference that a history of any type of treatment for amblyopia did not exclude participation. In brief, amblyopic patients were eligible to participate if their fully corrected vision (following a period of optical treatment, usually of minimum 16 weeks, and showing no improvement in acuity for at least the last 2 consecutive visits) was characterised by at least 2 logMAR lines of inter-ocular difference in acuity, with the FE showing acuity of 0.2 logMAR or better. The type of amblyopia could be i) strabismic, in the presence of heterotropia at distance and/or near fixation or misalignment induced by hyperopic spectacle correction, ii) anisometropic, when there was an interocular difference of minimum 1.0 dioptre (D) in spherical equivalent or minimum 1.50D in astigmatism at any meridian or iii) combined, when criteria i) and ii) were both satisfied. The control group of children (n=10) were initially selected based on their age (3-12 yrs).

The general exclusion criteria were i) a history of other ocular pathologies or physical forms of deprivation (e.g. cataracts, ptosis, macular degeneration) that might have caused amblyopia and/or further complications and ii) a diagnosis of significant neurological deficit (dyslexia, autism, epilepsy etc.) or iii) other neurological complications. Participants with 'intermittent strabismus' (where the degree of eye-misalignment varies over time) were also excluded.
Henceforth we refer to the groups of 'Adults with Amblyopia', 'Children without Amblyopia' and 'Children with Amblyopia' respectively as '*A+A*', '*C-A*' and '*C+A*'. Participants' identification number for this study (ID), age, gender and their visual acuity (all from clinical records) are provided in
Table **9**.

Group	ID	Age (yrs)	f/m	Acuity LE	Acuity RE	IOAD	Amblyopia Type
C-A	1	9.7	f	0.10	0.06	0.04	-
	2	11	m	0.22	0.14	0.08	-
	3	8.58	f	0.06	0.00	0.06	-
	4	8.38	f	0.02	0.02	0.00	-
	5	8.02	m	0.16	0.12	0.04	-
	6	6.21	f	0.00	0.18	0.18	-
	7	8.08	m	-0.06	0.12	0.18	-
	8	5.53	m	0.02	0.00	0.02	-
	9	8.22	m	0.18	0.00	0.18	-
	10	7.3	m	0.14	0.22	0.08	-
C+A	11	6.94	f	0.52	-0.04	0.56	comb.
	12	11.1	m	0.00	0.30	0.30	comb.
	13	7.07	f	0.22	0.00	0.22	aniso.
	14	5.47	f	0.54	0.10	0.44	comb.
	15	6.45	m	-0.04	0.20	0.24	comb.
	16	8.01	f	***1.20***	0.02	1.18	aniso.
	17	7.73	m	0.46	0.20	0.26	strab.
	18	7.09	m	0.00	0.22	0.22	strab.
	19	10.3	f	0.10	0.40	0.30	strab.
	20	5.99	f	0.14	0.44	0.30	aniso.
A+A	1	61.53	m	0.3	-0.06	0.36	comb.
	2	25.08	f	0.10	***1.78***	1.68	comb.
	3	30.31	f	0.08	***0.70***	0.62	comb.
	4	60.33	f	0.00	0.50	0.50	comb.
	5	20.18	m	-0.10	0.12	0.22	comb.
	6	43.02	m	-0.12	0.18	0.30	strab.
	7	53.85	f	0.00	0.22	0.22	aniso.
	8	47.91	f	0.18	0.48	0.30	strab.
	9	32.79	m	0.48	0.08	0.40	strab.
	10	38.09	m	0.00	0.30	0.30	comb.

Table 9 Details of participants, subdivided by group (indicated in column 1): C-A, children without amblyopia, C+A children with amblyopia and A+A, adults with amblyopia. Participants' ID numbers are given in column 2. The best corrected visual acuity, measured using crowded line-charts, is reported in LogMAR for both the left-eye (LE; column 5) and the right-eye (RE; column 6); the inter-ocular difference in acuity (IOAD) is also indicated, for the reader to easily check for the presence of amblyopia (if IOAD≥0.2 logMAR). When applicable, grey-shaded cells highlight the acuity in the amblyopic eye and severe cases of amblyopia are indicated by underlined values (acuity in the amblyopic eye >0.6 LogMAR). When present, the type of amblyopia is reported in the last column, as 'combined mechanism' (comb.; N=10), 'strabismic' (strab.; N=7) or 'anisometropic' (aniso.; N=3) amblyopia.

Amblyopic participants: details

Children and adults with amblyopia were classified by sub-type (

Table 9, last column). Within group C+A, children were classified as anisometropic (n=3), strabismic (n=3) or combined-mechanism (n=4). Only one child had severe amblyopia (acuity in the AE >0.6 logMAR; type: anisometropic) while three showed mild amblyopia (acuity in the AE<0.3logMAR; respectively, with anisometropic, combined and strabismic amblyopia). Within the group of adults, one had anisometropic amblyopia, three were strabismic and six had combined-mechanism amblyopia. In this group, two had severe amblyopia (both, combined mechanism) while one had mild amblyopia (strabismic and combined-mechanism). So, overall, 11/20 patients (including six children) showed moderate amblyopia (i.e. baseline best-corrected acuity in the AE between 0.3 and 0.6 logMAR, included). In total, 7/20 participants with amblyopia (including five children) showed a left-eye deficit, so that we would expect the other 13/20 patients to show a left-eye dominance (having them a right-amblyopia). This result was confirmed, as visible in Figure 39.

Participants: sample size

Ten participants per group is a not a large cohort, although we note that other studies in adults (e.g. M. Piano & Newsham, 2015) and children (e.g. N. Herbison et al., 2013) have been conducted with similar sample sizes. However, the aim of the pilot study reported in this chapter was to further investigate the tests introduced in Chapter 5, to identify those tests that were both reliable in detecting SED in a clinical population and accurate in differentiating participants, based on the presence or absence of amblyopia. We designed this study based on the minimum sample size for each group (three in total; C-A, C+A, A+A) needed to detect, with a 95% confidence level and 80% power, whether a stated difference exists between the mean-BP measured in participants with and without-amblyopia. We based our sample size calculation on conservative performance-estimates derived from our previous studies on children with amblyopia (Chapter 3) and adults without amblyopia (Chapter 5). As described in Chapter 3, the mean-BP obtained in children with amblyopia was 0.71 (averaging between the 22 participants the individual estimates, derived from the average BP acquired daily, for the duration of BBV treatment) and the associated variance was 0.02 (derived from the σ=0.13 of the series of individual mean BPs, obtained from the 22 children who completed BBV treatment; specifically, each mean-BP was calculated by averaging the individual daily estimates, from the 'ghost-task'). In Chapter 5, we reported that the mean-BP in adults with normal (or corrected to normal) vision was 0.52 (variance was 0.001). We considered the variance in children as an indication of the general population

(we assumed a more 'noisy' scenario, as children usually demonstrate a higher variability in performance than adults).

The recommended minimum sample size for each group was N=9. This number was calculated using the formula:

$$n = (Z_{\frac{\alpha}{2}} + Z_{\beta})^2 * 2\frac{\sigma^2}{d^2}$$, for α=0.05

where $Z_{\frac{\alpha}{2}}$ is equal to 1.96 (the critical value of the Normal distribution for $\frac{\alpha}{2}$=0.025) and Z_{β} is 0.84 (where β=1- 'desired power' =0.2, and 0.84 the critical value of the Normal distribution for β), σ^2=0.02 (the considered variance); d, i.e. the difference between the mean-BP in the two populations, was equal to 0.19. We slightly raised the sample size in each group to N=10 to be more likely to ensure the recommended minimum of 9 participants per group, considering possible outliers (and/or unlikely withdrawals).

Note that in Chapter 3, SED=1 was indicative of full reliance on the *fellow*-eye, while in chapter 5, it was indicative of full reliance on the *right*-eye. However, in our knowledge only three adults without-amblyopia (see Chapter 5) demonstrated an inter-ocular difference in acuity (with ≤0.2 logMAR acuity in either eye). So, looking at the *amount* of inter-ocular difference in SED, the fact that we measured the Balance-Point on a scale of left/right eye instead of non-dominant/dominant eye made no difference in adults without amblyopia (whose SED was almost balanced). Therefore, the BP indexes obtained in adults (without amblyopia) are comparable with those recorded in amblyopic children and adults on the scale AE(non-dominant) / FE(dominant)-eye.

6.2.2 Apparatus

Stimuli were presented on a linearised LG 23" LCD 3D monitor, with a 1920x1080 pixel resolution operating at 120Hz. Stimuli were generated in Matlab (Mathworks Ltd) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007). During testing, stimuli were viewed through wireless LCD shutter glasses (nVidia Corp., Santa Clara, CA) allowing independent control of the image presented to each eye (so that effective framerate was 60Hz per eye). Shutter glasses were worn over optical correction, when necessary, and, for children, were mounted in a customized ski mask to ensure comfort while maintaining a snug fit over spectacles. A Minolta LS110 photometer (Konica-Minolta Ltd) was used to calibrate the monitor-luminance, using measurements made through the shutter glasses. Participants viewed the screen at a distance of 100 cm to produce a pixel density of 66.3 pixel/deg. Participants communicated their responses verbally. The experimenter recorded responses using a dedicated keyboard.

6.2.3 Stimuli, Procedure & Analyses

Stimuli, procedure and analyses are fully described in section 5.2.3. In brief, tests #1-7 estimated observers' Sensory Eye-Dominance (SED) in term of a Balance-Point (BP) that quantifies how much the signal received by each eye, under dichoptic viewing, contributes to performance of a task. Results were clamped in the range 0.0-1.0, indicating the proportion of left- versus right- eye predominance (where 0.0=exclusive left-SED; 1.0=exclusive right-SED). Specifically, tests #1-6 estimated observers' *contrast-BP*, measuring the ratio between the contrast-level applied to each element of paired-stimuli to make each one being equally likely to be chosen, with 50% identification. Testing consisted of having participants indicate (based on a two-alternative choice):

- the 'whitest' letter, chosen from pairs of either non-overlapping letters (test #1), overlapping letters (test #2) or polarity-rivalrous letters (test #3)
- vertical position (above or below), compared to a horizontal reference line. Stimuli were composed of non-rivalrous (test #4) or rivalrous (test #5) phase-shifted-gratings
- perceived direction (up versus down) of rivalrous noise-patterns, moving in opposite directions (test #6).

Unlike other measures, test #7 estimated BP as the proportion of frames when the participant's response was consistent with his/her relying on the left-eye view to decide which moving stimulus dominated perception during *binocular-rivalry.* Thus, for this test the BP quantifies the *duration* of motion-direction dominance (using the same type of stimuli as in test #6). In addition to the SED tests, test #8 measured *stereoacuity*: the minimum detectable horizontal-disparity between stereo-defined dichoptic letters. All tests were performed twice in two separate runs.

We assessed tests according to their 1) *reliability* (consistency of two measurements from the same test performed on a given participant), and 2) *validity* (the ability of the test to capture individual differences amongst participants). Reliability, as the change in performance between run 1 and 2, was quantified using two metrics. First, the Bland Altman 95% limits of agreement, or Coefficient of Repeatability (CoR) – with lower values indicating higher repeatability (Vaz et al., 2013). Second, the Intraclass Correlation Coefficient (ICC; based on a mean rating-k=2, absolute agreement, 1-way random-effects model) – with higher values indicating higher repeatability (McGraw & Wong, 1996). We also calculated two other metrics that incorporate both reliability and validity. First, the Mean-Average-Precision (MAP; a

statistic that is less sensitive to outliers than the previous metrics). Second, Fractional-Rank-Precision (FRP; a statistic that quantifies how identifiable a participant is from their set of the test-scores). Higher MAP and FRP values indicate higher reliability and validity.

6.3 Results

We present the data collected on 'Adults with Amblyopia' (group A-A, N=10), 'Children without Amblyopia' (group C-A, N=10) and 'Children with Amblyopia' (group C+A, N=10). We also integrate to these results those obtained on 'Adults without Amblyopia' (discussed in Chapter 5; now labelled, group 'A-A', N=30).

Summary We asked how amblyopia would influence the results obtained on participants of a similar age, comparing C-A to C+A and A-A to A+A groups. Figure 38 (top) shows how the individual values of mean magnitude of sensory-eye-dominance (SED-magnitude) resulted similar for participants within the same group but, overall, differed from a group's participants to another's. This pattern is evident both for adult-groups (Figure 38, left graph) and children-groups (Figure 38, right graph). The individual mean SED-magnitude, expressed in %, was calculated as

$$\left[\left(\frac{\sum_{i=1}^{n}\left|\left(\frac{\mathrm{BP}_{run1}-\mathrm{BP}_{run2}}{2}\right)_i-0.5\right|}{n}*100\right)+50\right],$$

where *i*=test # and *n*=7. Figure 38 reports the resulting values, from 50%=balanced vision (equal preference for each eye, i.e. no eye dominance) to 100%=fully unbalanced vision (complete preference for one eye, i.e. the fully dominant-eye in binocular vision). Henceforth, we will refer to the individual BP resulting from the mean between runs (i.e. the term $\frac{\mathrm{BP}_{run1}-\mathrm{BP}_{run2}}{2}$ in the above equation) simply as BP.

To check for significant differences between participants of a similar age, we ran a two-sample t-test for each pair of groups (of adults, and of children), comparing the series of SED-magnitude respectively obtained from participants with- and without- amblyopia. For the group A-A (N=30), we pooled one-every-three BPs (BPs in ascending order) to get a series of 10 estimates, i.e. as many as in the other three groups (pooled group labelled: 'A-Ap'). This allowed us to have the same degree of freedom (d.f.=9) between the t-distribution of each group (thus not affecting the probability of rejecting the null hypothesis, i.e. there is no difference between groups in term of SED-magnitude). The t-test analysis revealed a statistically significant difference (*p*=0.0046) between C+A and

C-A, as well as a significant difference (p=0.0001) between A+A and A-A. The presence of amblyopia significantly affected the mean SED-magnitude regardless of participants' age (i.e. if adults or children), as a 2-sample t-test between the 20 amblyopes (C+A and A+A) and 19 non-amblyopes (C-A, excluding outlier ID1, and A-Ap) also gave a significant result ($t_{37d.f.}$=-4.716, p<0.001). For each group, the mean and standard deviation of the SED-magnitude measures were: μ=0.13, σ=0.036 for C+A, μ=0.06, σ=0.048 for C-A (excluding ID1, outlier), and μ=0.29, σ=0.164 for A+A, μ=0.03, σ=0.029 for A-A (including the data from all 30 participants). As expected, children and adults with amblyopia demonstrated greater SED imbalance than non-amblyopic participants. Note that among adults, only participants assigned to A+A were patients, while those in A-A were recruited within the university and therefore were not undergoing a standard clinical assessment of their vision. In contrast, all children, assigned to either C+A or C-A, were recruited at the hospital, indicating the likely presence of a problem with their vision (see our inclusion and exclusion criteria for more details). Remember that we excluded children having ptosis, nerve palsy or other conditions linked to deprivation-amblyopia but we did *not* exclude those showing no-stereopsis at standard clinical tests. As a possible result, the magnitude of SED for C-A (μ=0.10, σ=0.11 when including also ID1) was slightly higher than the one obtained for A-A; but this difference was not significant (considering A-Ap: two sample t-test: $t_{d.f.17}$=-1.63, p=0.122, excluding ID1 of C-A; $t_{d.f.18}$=-1.70, p=0.107, including this participant).

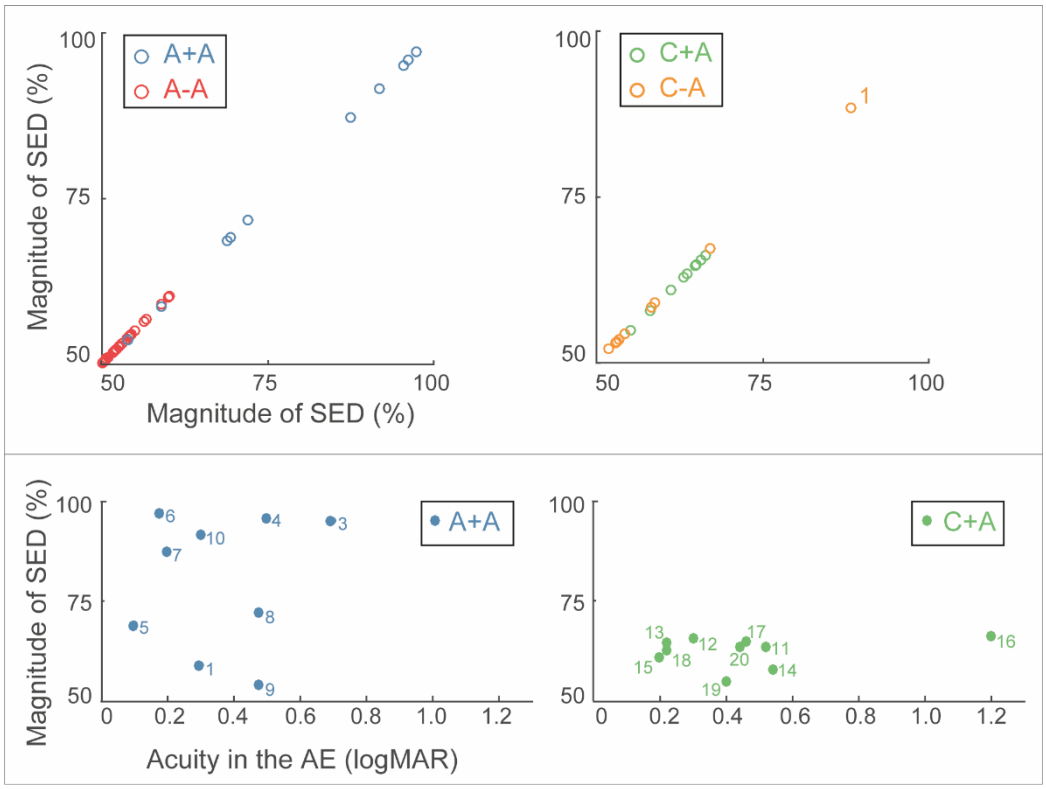


Figure 38 SED and acuity in A+A vs A-A and C+A vs C-A. **(Top)** plots the individual mean magnitude of Sensory-Eye-Dominance, expressed in percentage (|mean BP-0.5|*100). On the y-axis: 50=balanced vision, 100=complete eye dominance. The mean BPs, obtained by averaging the Balance-Points estimated in tests #1-7, are colour-coded per group. As specified in the legend, on the left are data from A+A and A-A, while on the right, are data from C+A and C-A. Note the clustering of the data per group (except for only a few; in particular, outlier ID1 from C-A, labelled). **(Bottom)** Magnitude of SED in participants with amblyopia, either adults (on the left; A+A) or children (on the right; C+A), plotted against the individual level of visual acuity in the AE (measured in logMAR). A clear relation between poorer acuity and higher SED imbalance does not appear.

Is the magnitude of SED related to acuity?

For each participant, we collected monocular-acuity estimates from individual clinical files. Among children without amblyopia (C-A), the best acuity between the two monocular results (i.e. the acuity in the 'fellow-eye', for comparison with C+A) resulted in between -0.06 and 0.14 logMAR (μ=0.04, σ=0.07); while, in the weaker-eye, measures ranged from 0.02 to 0.22 logMAR (μ=0.13, σ=0.08). Overall, the mean inter-ocular difference in visual acuity (IOAD) recorded in C-A was 0.09±0.07 logMAR and the individual scores varied from nil to 0.18 logMAR. In comparison, the range of variation of the individual SED-magnitude (mean SED across tests #1-7) was wider, ranging from 2% to 38%. For C-A participants, a correlation analysis between the individual IOAD and SED-magnitude revealed the absence of a significant relation (either if including ID1: R=-0.42, p=0.23, or excluding this participant: R=-0.50, p=0.16, who was considered outlier – see Figure 38-top).

On the other hand, A+A and C+A (Figure 38-lower box), i.e. participants diagnosed with amblyopia, respectively showed acuities in the FE within -0.12 to 0.18 logMAR and -0.04 to 0.2 logMAR. Considering the AE, the range of acuities went from 0.1 to 1.78 logMAR in adults (A+A) and from 0.2 to 1.2 logMAR in children (C+A). As expected, on average, the individual IOADs were greater than in C-A: we recorded IOAD values varying from 0.20 to 1.68 logMAR in A+A (μ=0.49, σ=0.44 logMAR), and from 0.22 to 1.18 logMAR in C+A (μ=0.40, σ=0.29 logMAR). Using a multilinear regression, at a 5% confidence level, we found that the magnitude of binocular imbalance could *not* be predicted by the individual level of IOAD, neither for A+A (R^2=0.020, *p*=0.70) nor for C+A (R^2=0.082, *p*=0.42).

Remember that one of our (widely used) criteria for amblyopia is that FE acuity should be better or equal to 0.2 logMAR. Therefore, poorer acuity in the AE would be associated to greater IOAD. Expecting similar results than to the analysis involving IOAD, we checked whether there was a correlation between the individual acuity in the AE and the magnitude of SED. Figure 38-bottom conveys how, especially in adults (graph on the left), there is a dispersion of individual SED-magnitudes also for those participants whose acuities in the AE were similar. It also shows, especially for children (graph on the right), how participants whose acuity in the AE were not reciprocally similar, obtained similar SED-magnitude results. Statistically, on average there was no correlation between the individual acuity in the AE and the SED-magnitude (*R*=-0.17, *p*=0.64, for A+A; *R*=0.24, *p*=0.51, for C+A).

Note that only 2/10 adults and 1/10 children with amblyopia showed a severe deficit (i.e. acuity in the AE poorer than 0.6logMAR; see
Table **9**). Therefore, the range of IOAD values that we considered, which were obtained mainly from patients with mild to moderate amblyopia, was relatively 'limited' and, possibly, not wide enough to reveal a significant impact on SED measures. Whether this was the case or not, the results suggest that there is a general unclear relation between acuity (a measure of monocular vision) and the *amount* of SED (a measure of binocular vision) and therefore both measures could independently contribute to a thorough vision assessment.

Stereoacuity results. Nine of the twenty children (four from C+A) and none of our amblyopic adults (A+A) completed two runs of our stereoacuity test (#8). For these nine children (indicated by an '*' next to the relevant observer IDs, in Figure 39), considered together, the statistical metrics resulted in: ICC=0.21, F=1.53 (*p*=0.27), CoR>1.00,

MAP=0.43 and FRP=0.74. These values suggested the results from this test are reliable, when looking at the inter-observer differences in relation to the variability of the sample (FRP) but less so when considering the 'absolute' reliability (ICC) and, in particular, the individual consistency in performance (CoR). However, the precision of repeated measurements was strong (MAP), considered the same index obtained in tests #1-7. The F statistic signalled only a moderate ability in differentiating participants, and we note this was biased by the huge inter-individual differences, mainly between C-A (n=4) and C+A (n=5). Overall, test #8 appears a potentially valid index of stereoacuity, but as we only managed to record results in children (and not all of them) we were not able to compare our SED-tests to this stereo test. Note that we report stereoacuity for four children (1 C+A) whose Frisby results were missed, despite our stereo-test including 20 trials instead of few isolated answers. However, we missed to obtain a result from other three children (1 C+A), who instead demonstrated measurable stereopsis at Frisby. This give an example of how relying wholly on stereoacuity when evaluating binocularity can be limiting. With this in mind, we consider what could have justified why test #8 was only feasible on 9/30 participants.

This test was always the last to be administered, to maximise the probability we would collect a full set of SED measures for all the participants. That less than 50% of children completed the two runs arose from a general inability to attend to the stereoacuity task for sufficient time. This is likely an issue during standard visual assessment in the clinic: we found a record of results from the standard clinical stereo-test (Frisby) for only 4/10 children without amblyopia and 7/10 with amblyopia. It is likely that for children without amblyopia, stereo-acuity is not routinely measured due to the limited time available during clinical appointments and the assumption that normal behaviour will be observed in most cases. In contrast, measurement of stereo-acuity is recommended for children with amblyopia: 70% of the children recruited had a record of Frisby test results (and/or other measures), although these measurements were inconclusive in 2 cases. We speculate that in the remaining 30% of cases, tests were not performed because the clinician's opinion (based on previous experience) was that the test would not produce a reliable estimate of stereopsis. It is unclear why none of the adults demonstrated measurable stereoacuity; we consider possible implications in Discussion (section 6.4).

How much do participants rely on their dominant eye? The left side of Figure 39, shows the mean balance points (here, averaged between the two runs) for each group, obtained for each participant on tests #1-7. Note that the data from some tests (e.g. 1,

2, 3) are less visible than other points (e.g. 4, 6, 7) because of the order points were plotted in (from #1 to #7).

Results from the 10 adults with amblyopia (A+A; top), 10 children without amblyopia (C-A; middle graph) and 10 children with amblyopia (C+A; lower graph) are shown. Data are ordered in ascending order of participants' individual overall mean-BP (averaged across tests #1-7). Participants' correspondent IDs are reported on the x-axis. The overall mean-BPs were in the range [0.03-0.87] for group A+A, [0.12-0.53] for C-A and [0.34-0.66] for C+A. Respectively, the associated standard deviations were 0.278 (for A+A), 0.124 (0.07 excluding ID1; for C-A) and 0.138 (for C+A). We checked whether the variance in the series of mean-BPs was significantly different between participants with- and without- amblyopia, as one would expect, regardless of their age. An *F*-test between 20 amblyopes (C+A, A+A) and 19 non-amblyopes (10 A-Ap; 9 C-A, excluding ID1) indicated that there was significant evidence that the two series of data came from populations with *un*-equal variances ($F_{\text{d.f.1=19, df2=18}}$=14.261, p<0.001). In contrast, one would expect samples pooled from the same population to show no or non-significant difference between the mean variance in the respective estimates (here, the mean-BPs between either +A or -A groups). In fact, the F-test between A-Ap and C-A (excluding ID1) revealed a *non*-significant difference ($F_{d.f.1=9,\ \text{d.f.2=8}}$=0.436, p=0.239). However, we found a marginally significant difference between the variances of mean-BPs obtained in A+A and those obtained in C+A ($F_{\text{d.f.1=9, d.f.2=9}}$=4.044, p=0.049), possibly as a consequence of the concomitant conditions that might have cumulated with aging (note, in Figure 38, the greater SED imbalance demonstrated by A+A compared to C+A). The results indicate a prevalence of left-eye dominance in 7/10 children without amblyopia. Among the amblyopic participants, a left-eye preference was demonstrated in 8/10 adults and 5/5 children, as expected based on the side of their AE (see Table **9**).

The right side of Figure 39, plots the BP-index from run 1 (blue symbols) and run 2 (red symbols) from two participants from each group. These participants were chosen based on their ranked overall-mean BP score. The first plot (on the right side of in each box) is from the participant with the second-highest ranking for left-eye preference, the second plot is from the participant with the second-highest ranking for right-eye preference. We note a clustering of measures from the two runs, although the error bars from Palamedes refits are sometimes large. Nonetheless, the green lines in each of these plots (which show the mean BP for that participant, also shown as the green-boxed numbers and as green-triangles, for each participant, on summary graphs on the left side of Figure 39), indicate a high level of agreement across runs of the same test. This agrees with results

from Chapter 5. We computed the mean difference between participants in the average score between runs, obtained across tests #1-7, and found a good general agreement – the mean difference was μ=0.01 (σ=0.05) in A+A, μ=0.05 (σ=0.08) in C-A and μ=0.03 (σ=0.10) in C+A. Even in cases where the error bars were large, the estimated BP-index does not diverge much from the overall-mean BP. On top of the strong clustering of results from runs 1 and 2, we note the same for results across tests (averaging, for each test, the difference across participants): μ=0.01 (σ=0.02) for A+A, μ=0.05 (σ=0.06) C-A and μ=0.03 (σ=0.04) in C+A.

We examined the consistency of contrast-balance estimates using a correlation analysis. We observe consistency between contrast-balance estimates from different tests, as indicated by the mean correlation coefficient (R= 0.84 ±0.14; obtained by averaging the R resulting across tests #1-6) and the associated *p*-value (*p*=0.0024). Specifically, for each test, we calculated the correlation between the individual BPs obtained in that test and the individual mean-BPs, across the other five tests. Only test #6 showed a non-significant correlation (R 0.562, p=0.091), indicating a generalised different performance in this test compared to the one demonstrated in the other tests.

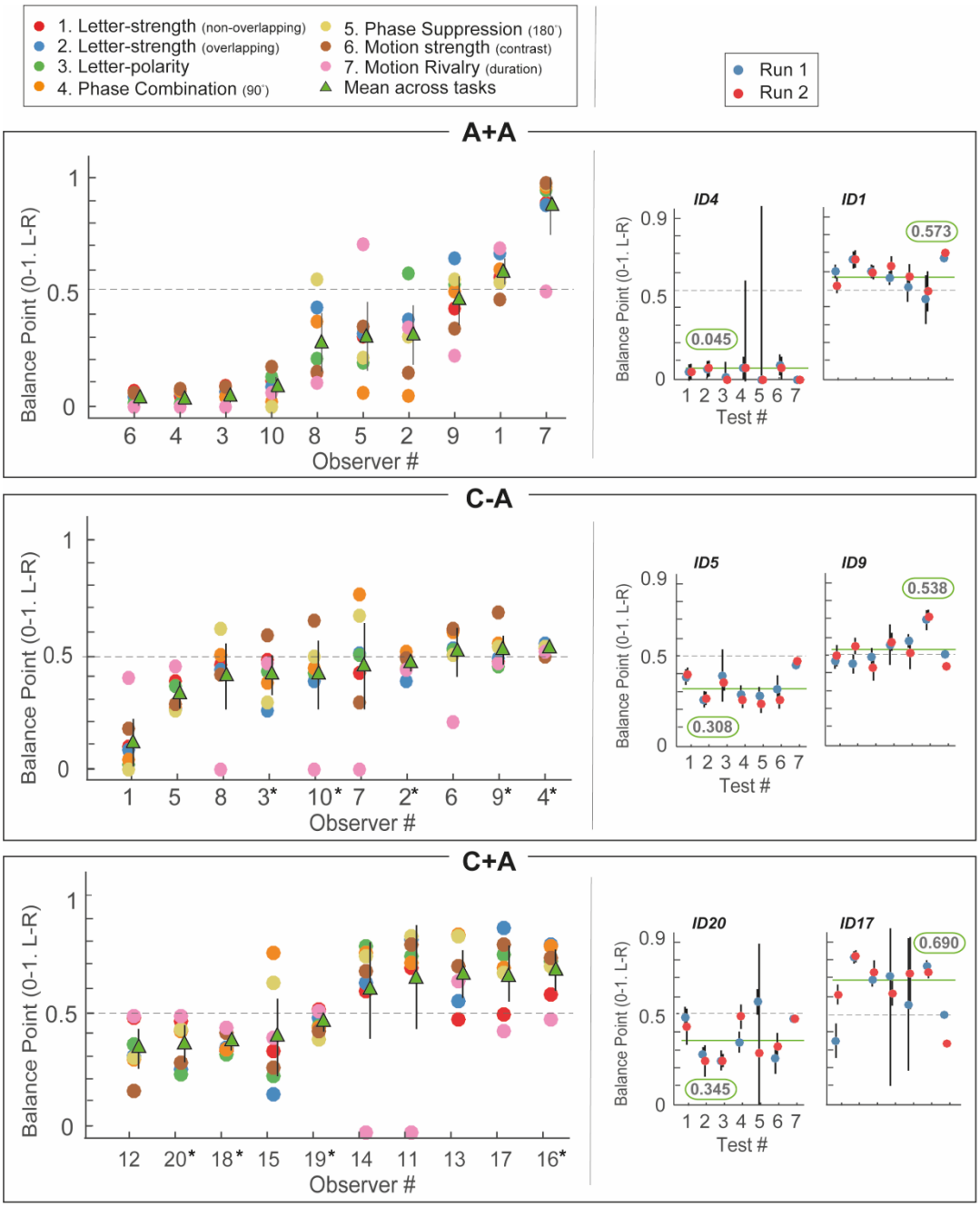


Figure 39 (left side) Estimated balance point (BP) for all participants from group A+A (top), C-A (middle) and C+A (bottom). Participants within each plot are ordered by mean BP across tests (green triangles; error bars are 95% confidence intervals). Next to the relevant IDs, a '*' indicates those children who completed two runs of our stereoacuity test (#8) – none of the adults did. As in Figure 34, a BP of 0 indicates complete reliance on the left eye, 1 complete reliance on the right eye and 0.5, perfect inter-ocular balance. Tests are colour coded according to the legend. **(right side)** Illustrative results from two participants from each group. Blue and red symbols show data from the first and second run respectively, and error bars are 95% confidence intervals on estimates. Mean BP across tasks (green line) is the green-boxed figure in each panel.

Which is the "best" test? Figure 40, Figure 41 and Figure 42 are Bland-Altman plots of the data from the seven SED tests, including the six contrast-BP tests (#1-6) and the rivalry test (#7). These plots show each test's ability to capture variability across participants (test-*validity*; high validity is associated with wider horizontal-spread of data) and the repeatability of measurements (test-retest *reliability*; higher reliability is associated with a narrower vertical spread of data, plotted as the 95% confidence intervals across estimates). Like Figure 35, a single data point corresponds to the individual threshold measured on that specific test and, in each Figure, there are 7 sub-plots, one for each SED test (labelled with test # at the top of each sub-plot). Data points are colour coded as belonging to the adult amblyopes group (A+A; Figure 40) or to one of the two groups of children (C+A, Figure 41 and C-A, Figure 42).

Looking at Figure 40 and Figure 41, performance in the presence of amblyopia does not seem to follow a specific pattern. Some tests elicit a wide range of balance estimates, notably the phase-tests (#4, #5) in adults and the two contrast-letter discrimination tests (#1, #2) in children. In particular, results from test #1 varied greatly across runs for participants from C+A, which exhibited good test-retest reliability in the other tests, compared to all other groups (including the A-A). Compared to C-A, the difference in the performance on this test cannot be explained by differences in age, which was, on average, similar (μ=8.04, σ=1.51 in C-A and μ=7.62, σ=1.80 in C+A). Recall that tests were always administered in the order #1 to #7, so children with amblyopia might have been 'overly-challenged' in this first test, due to their poorer baseline vision. Figure 42 (C-A) shows that tests #4 and 5 elicit wide variability in performance across participants. As was the case for group A-A (Figure 35, Chapter 5), test #1 yields the best repeatability in C-A, but (excluding participant ID1, previously indicated as outlier, as visible here too) this test yields a limited range (only ±0.1) of estimates of BP around 0.5. This test (#1) is insensitive to individual differences in performance compared to e.g. test #2. Only when tests #1 and #3 were run on group C-A did they exhibit similar performance. When these tests were administered to the other two groups of participants only test #3 led to a good compromise between test-retest reliability and test-validity.

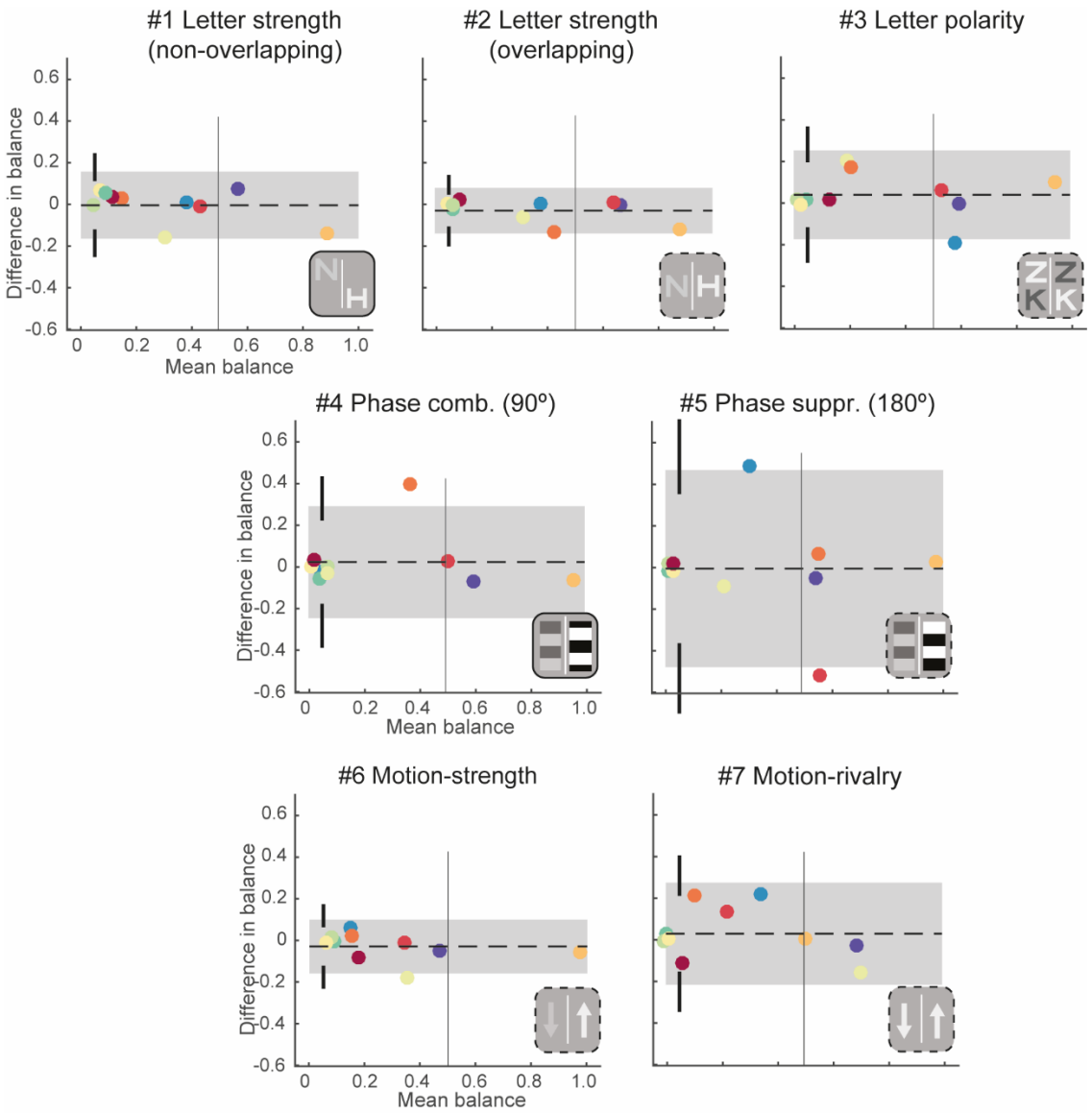


Figure 40 Bland-Altman plots of results from tests #1-7, obtained from *adults with amblyopia* (A+A). (#1-6) Mean contrast balance, averaged across runs, is plotted against the difference in contrast balance across runs. The plot of data from test #7 indicates the relative time spent in each rivalrous state plotted against the difference in this measure across runs. Shaded regions denote 95% confidence intervals across estimates with error bars on these estimates calculated using the procedure described in Carkeet (2015).

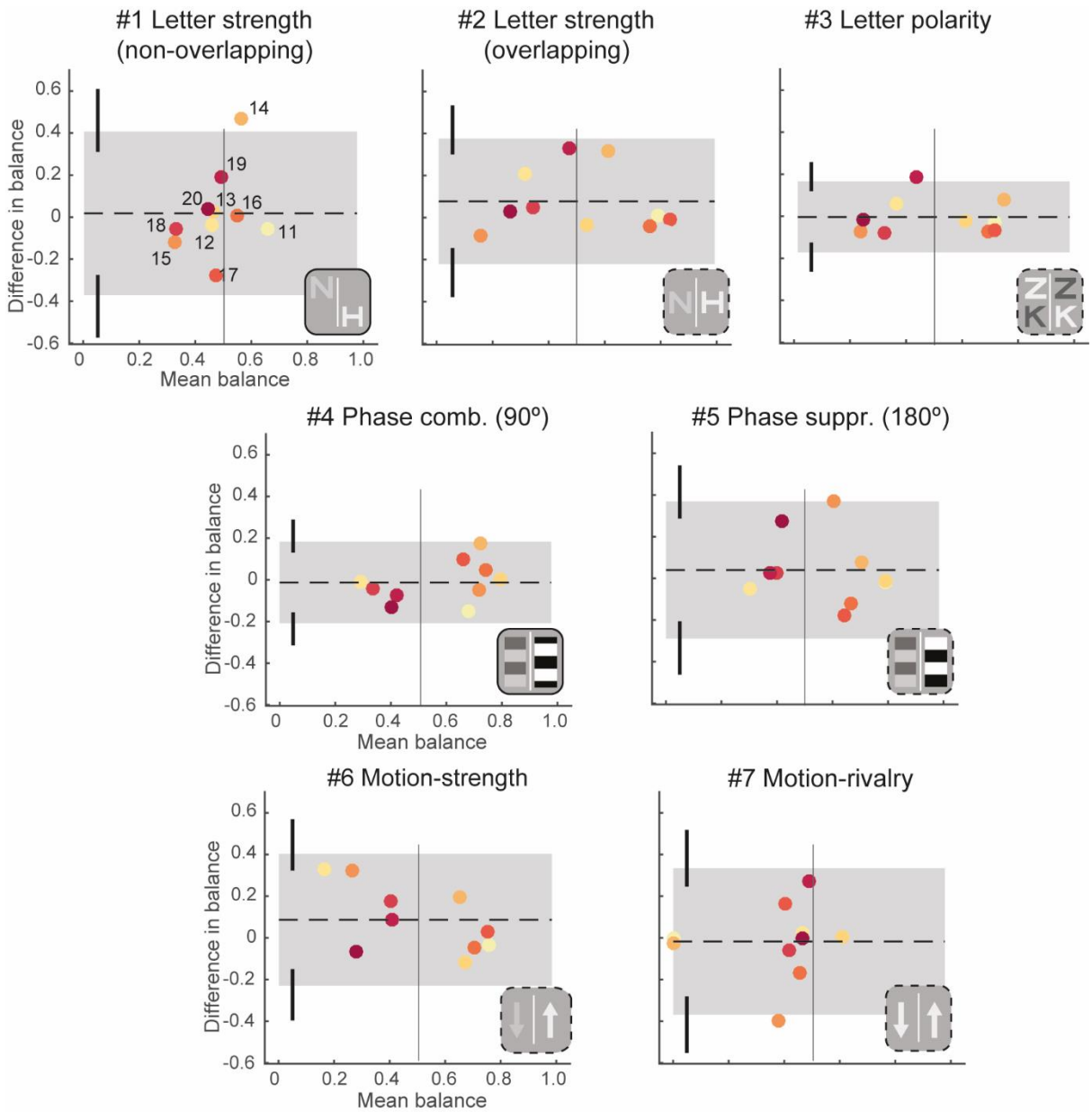


Figure 41 Bland-Altman plots of results from tests #1-7, obtained from *children with amblyopia (C+A).* Plotting conventions are as for Figure 40. Colours are all warm, in contrast to Figure 42, to indicate the presence of amblyopia. Different hues from children ID11 to 20 are labelled with ID numbers in the first sub-plot.

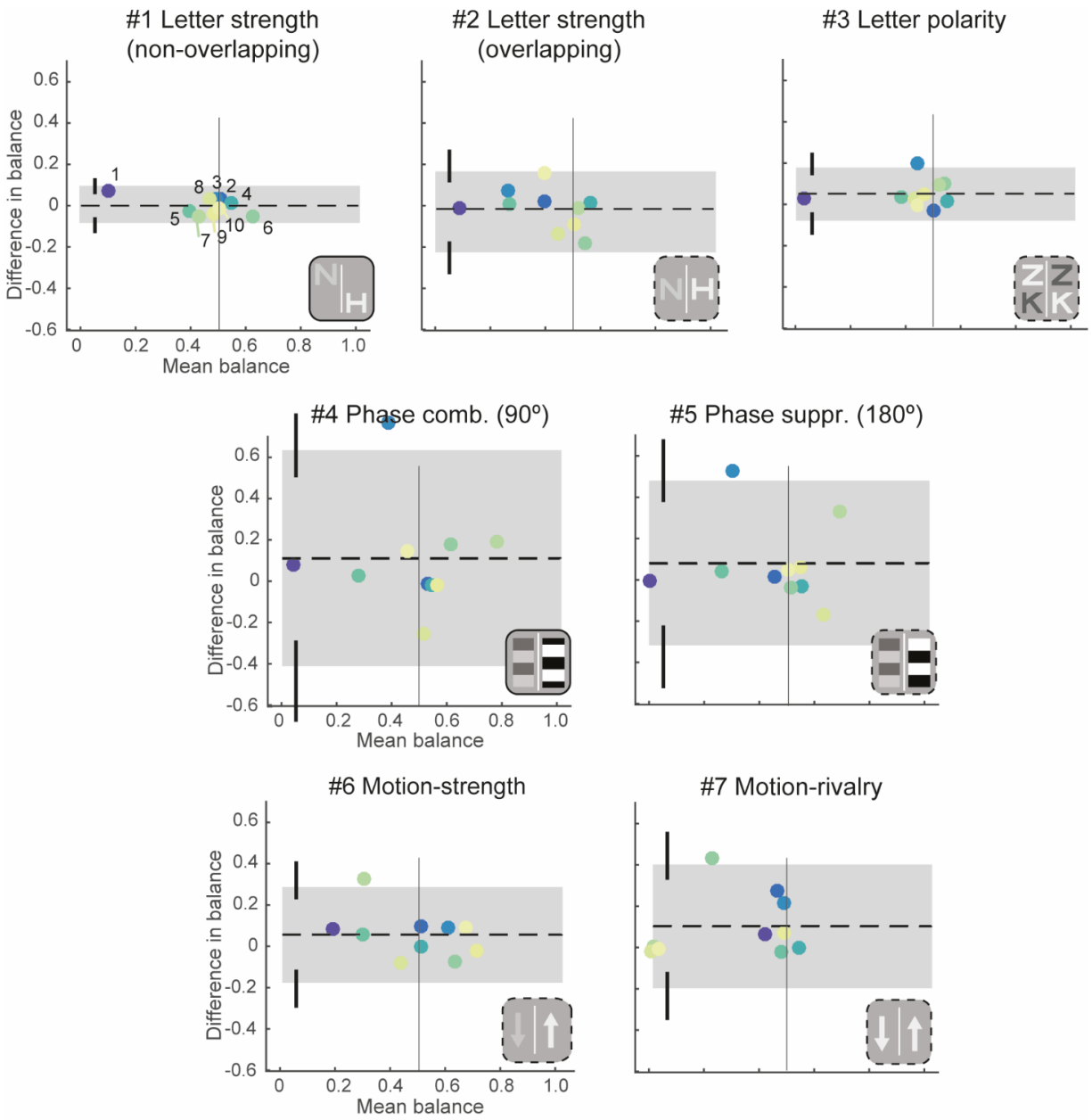


Figure 42 Bland-Altman plots of results from tests #1-7, obtained from *children without amblyopia (C-A).* Plotting conventions are as for Figure 36. Colours are cold, in contrast to Figure 37, to indicate the absence of amblyopia. Different hues from children ID1 to 10 are labelled with ID numbers in the first sub-plot.

A variety of statistics quantify the ability of a test to maximise the range of measures across participants (test-validity) while minimising the range of measures across repeated measurements (test-reliability; in our studies, across two runs). To visualise a comparison between the seven tests, the right column in Figure 43(c) shows one plot per group. Findings are broadly similar to what we found in A-A (previous chapter) both for A+A (top) and C-A (middle) but are somewhat different in C+A (bottom). Notably, while test #3 continued to perform well with children, test #1 did not.

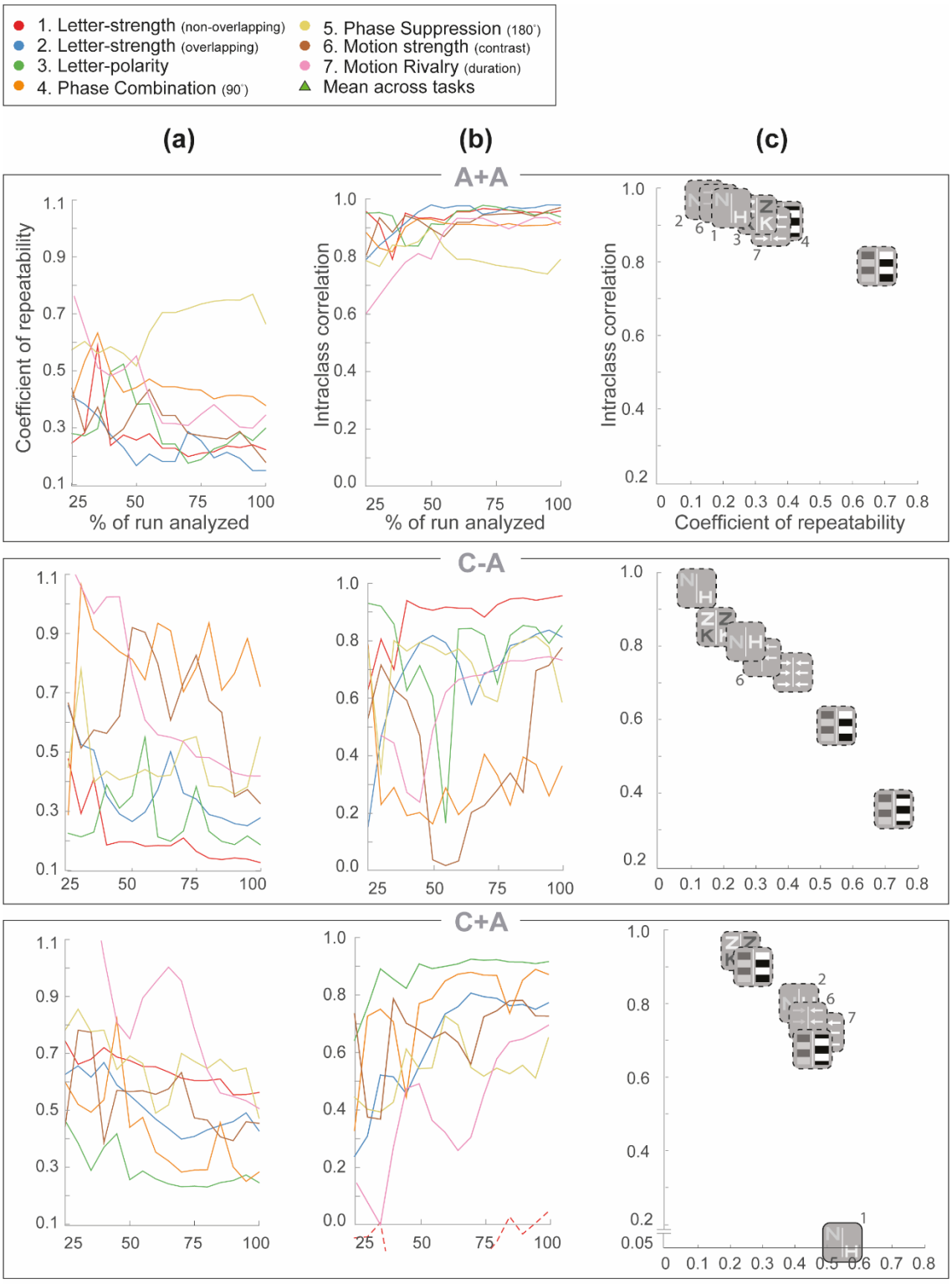


Figure 43 Statistics for our three groups (reported in each box: A+A: top, C-A: middle, C+A: bottom). There are three graphs in each box, forming three columns (a,b,c), labelled above the top box. In each graph, we plot respectively how: **(a)** coefficient of repeatability (lower is better) and **(b)** interclass correlation (higher is better) evolve over run-length (expressed as a percentage of the run analysed, e.g. 100%= all 20 trials) for the seven balance tests; **(c)** plots of intra-class correlation against coefficient of repeatability for all tests. The tests are listed in the legend-panel, in the top-left corner of the figure, and corresponding numbers are reported in (c) where required for clarity. Note that the y-axis in (c) for C+A has been modified to report the result for test#1.

Table 10 Table 12 summarise the statistics we applied to the results of the various tests applied to our three groups. Looking at ICCs and corresponding *F* values, tests #2, #1 and #3 appear particularly useful for groups A-A, C-A and C+A respectively. However, the ICC metric in children with amblyopia was substantially lower than in any other group. The Bland-Altman plot for test #1 (Figure 41) highlights this, as there is poor homogeneity between repeated measurements which translates into poor absolute-agreement in the variance between randomly chosen pairs of measurements, i.e. ICC<.5 (Koo & Li, 2016). ICCs are known to be disproportionately influenced by values at the ends of the measured range, making them susceptible to outliers. Rather than attempting to identify specific participants that might be responsible for the results, in order to circumvent this issue, we also measured MAP and the related FRP estimates which score on rank rather than absolute value. This difference in method impacts test #2 the most, in C+A, as it has a poor ICC (and *F*), but fair MAP and FRP values, but does not seem to have a significant impact on the ranking of other tests.

To rate each test's reliability, we looked more closely at both the *relative* performance between participants (ICC) and *absolute* performance in repeated measurements (CoR) (Vaz et al., 2013) and classified reliability adapting the guidelines suggested in Koo and Li (2016) to report ICC results. We note overall good reliability (0.75≤ICC<0.90) for tests #2 and #3, and moderate reliability (0.5≤ICC<0.75) for all other tests (Table 13). All tests run on amblyopic adults (A+A; Table 10) exhibited excellent reliability (ICC≥0.90) except test #5, which elicited moderately reliable performance in all other groups. Phase tasks were poorer than all other tests in group C-A (Table 11). The same was not evident in C+A (Table 12). The performance of tests with this group varied from moderate (#5, #6, #7), good (#2, #4) to excellent (#3). Test #1 scored an exceptionally low ICC index for the C+A group, driven by data from ID14 which showed very poor test-retest reliability. This reflected, to a lesser extent, the general trend in the results – note the dispersion of data points in Figure 41.

Tables 10, 11,12 and 13 - Summary statistics for the performance of the seven tests with each group (group identity is indicated above each table). (*) Indicates best performance (across tests #1-7) for a given metric. *ICC*: intraclass correlation, *F*: ANOVA *F* statistic, *CoR*: coefficient of repeatability (lower is better), *MAP*: mean average precision (higher is better), *FRP*: fractional rank precision (higher is better; Dorr et al., 2017)

Table 10 Statistics: Adults with Amblyopia

	ICC	*F (p)*	*CoR*	*MAP*	*FRP*
1. Letter-strength (non-overlapping)	0.959	48.321 (<0.00001)	0.227	0.687	0.890
2. Letter-strength (overlapping)	**0.979***	96.539* (<0.00001)	**0.154***	**0.808***	**0.940***
3. Letter-polarity (overlapping)	0.939	31.620 (<0.00001)	0.302	0.642	0.900
4. Phase-comb (90°)	0.921	24.266 (0.00001)	0.382	0.628	0.840
5. Phase-suppression (180°)	0.790	8.510 (0.00123)	0.669	0.486	0.780
6. Motion strength (contrast)	0.971	68.161 (<0.00001)	0.183	0.670	0.870
7. Motion rivalry (duration)	0.912	21.648 (<0.00002)	0.348	0.551	0.830

Table 11 Statistics: Children without amblyopia

	ICC	*F (p)*	*CoR*	*MAP*	*FRP*
1. Letter-strength (non-overlapping)	**0.957***	45.127* (<0.00001)	**0.117***	**0.565***	**0.810***
2. Letter-strength (overlapping)	0.813	9.171 (0.00071)	0.272	0.558	0.780
3. Letter-polarity (overlapping)	0.853	12.601 (0.00023)	0.179	0.464	0.730
4. Phase-comb (90°)	0.363	2.141 (0.12567)	0.730	0.491	0.710
5. Phase-suppression (180°)	0.587	3.847 (0.02359)	0.553	0.404	0.700
6. Motion strength (contrast)	0.775	7.904 (0.001667)	0.322	0.522	0.780
7. Motion rivalry (duration)	0.733	6.483 (0.00364)	0.418	0.376	0.740

Table 12 Statistics: Children with amblyopia

	ICC	*F (p)*	*CoR*	*MAP*	*FRP*
1. Letter-strength (non-overlapping)	0.052	1.110 (0.43313)	0.564	0.402	0.690
2. Letter-strength (overlapping)	0.780	8.098 (0.00151)	0.428	**0.601***	**0.820***
3. Letter-polarity (overlapping)	**0.924***	25.310* (<0.00001)	**0.245***	0.501	0.790
4. Phase-comb (90°)	0.879	15.597 (0.00009)	0.283	0.451	0.760
5. Phase-suppression (180°)	0.658	4.846 (0.10723)	0.473	0.417	0.730
6. Motion strength (contrast)	0.733	6.503 (0.00360)	0.455	0.491	0.760
7. Motion rivalry (duration)	0.702	5.711 (0.00590)	0.507	0.505	0.710

Table 13 Statistics: Overall (*across groups A-A, A+A, C-A and C+A*)

	ICC	*F (p)*	*CoR*	*MAP*	*FRP*
1. Letter-strength (non-overlapping)	0.643	24.648 0.10832)	0.241	0.455	0.757
2. Letter-strength (overlapping)	0.825	30.172* (0.00056)	0.269	**0.541***	0.797
3. Letter-polarity (overlapping)	**0.880***	19.671 (0.00006)	**0.208***	0.477	**0.799***
4. Phase-comb (90°)	0.598	10.900 (0.05751)	0.391	0.435	0.726
5. Phase-suppression (180°)	0.666	5.398 (0.00890)	0.455	0.376	0.730
6. Motion strength (contrast)	0.699	21.123 (0.011287)	0.364	0.486	0.791
7. Motion rivalry (duration)	0.637	8.836 (0.03641)	0.370	0.393	0.705

Figure 44 illustrates a summary of ICC and CoR statistics derived from the average performance of all participants, i.e. from groups A-A, A+A, C-A and C+A. in each test, #1 to #7. Figure 44a and b respectively report the average evolution of ICC and CoR statistics over the duration of each test (color-labelled as indicated in the legend, on the left of the figure). Figure 44c compares these two metrics for each test (tests are here represented by the corresponding icon, matched as in the Bland-Altman plots for each group). Overall, optotypes tests #2-3 (and #1, to a lesser extent) appeared 'the best' in term of reliability, while the phase tests (#4-5) and, slightly less, the motion-rivarly tests (motion-strenght, #6, and motion-duration, #7) resulted less reliable. Possible explanations are considered in Discussion.

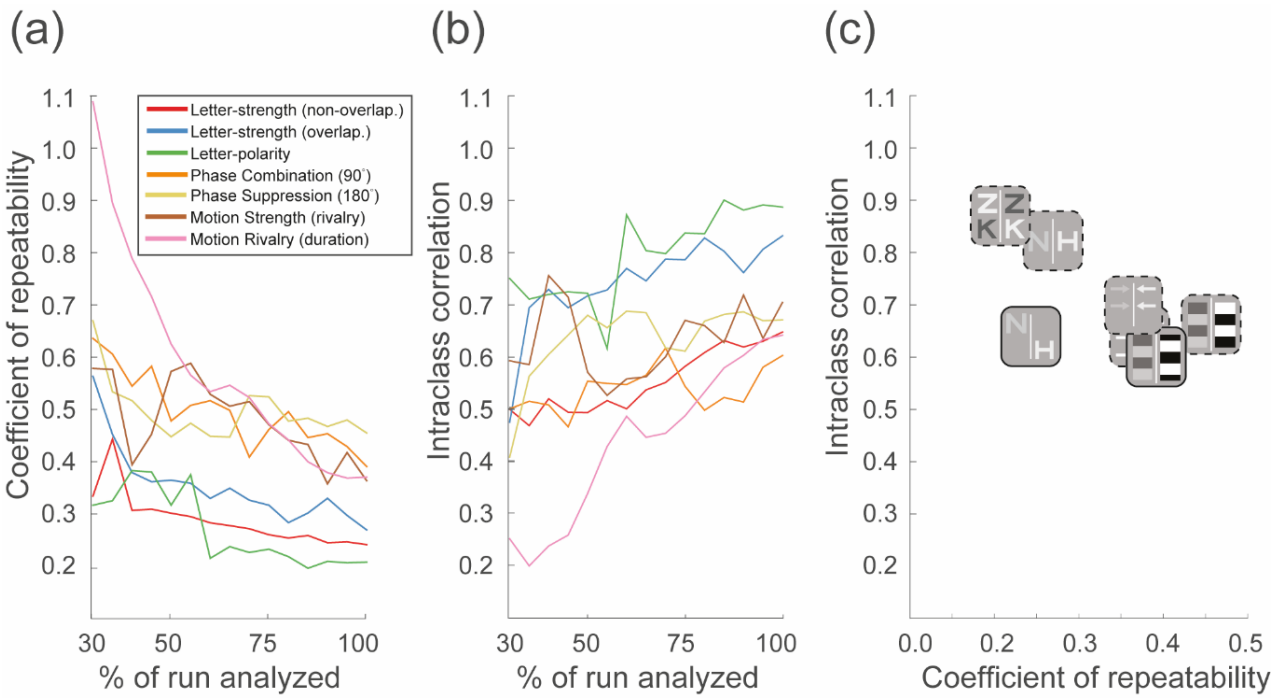


Figure 44 CoR and ICC across all 4 groups. Similar to Figures 33 (previous chapter, A-A) and 39 (A+A, C-A and CA). Plots indicate how (a) coefficient of repeatability (lower is better) and (b) intraclass correlation (higher is better) evolve over the 20 trials of a run. (c) a comparison between of these two metrics across tests. Data have been averaged across all four groups of participants.

6.4 Discussion

In this study, we have extended our investigation (described in Chapter 5) of a series of tests that quantify sensory-eye-dominance: on one hand, to children (20 in total) – to control for the influence of age, and on the other hand, to participants with amblyopia (20 in total; 10children and 10 adults) - to control whether the amblyopic disorder has a specific influence on performance. As in Chapter 5, we evaluated results in term of reliability (intra-observer test-retest) and test-validity (inter-observer agreement between measurements obtained on each test).

Summary of results. Whether comparing between each other the two groups of adults (A+A, A-A) or the two of children (C+A, C-A), the presence of amblyopia significantly influenced both the amount- and the distribution- of sensory-eye-dominance (SED). These measures were expressed as the individual mean, across tests #1 to #7, respectively of SED-magnitude (on a scale from 50% preference for either eye, to 100%/complete preference for one eye) and of mean-BP (on a scale from 0=left SED, to 1=right SED; from which SED-magnitude was derived). On average, we found a difference in SED-magnitude of +0.16 between participants without amblyopia (30 A-A plus 10 C-A; μ=0.05, σ=0.065) and those with amblyopia (10 A+A plus 10 C+A; μ=0.21, σ=0.146). As expected, the *side* of SED varied in accordance with the right- or the left-eye having a better visual acuity. So that the individual mean-BP fell, respectively, in the interval 0-0.5 (left-SED) or 0.5-1 (right-SED). However, for participants in groups C-A, A+A and C+A, the *amount* of SED (SED-magnitude) was <u>not</u> significantly correlated either with the level of monocular-acuity or with the interocular-difference-in-acuity. Thus, independent of acuity, SED represents another aspect of visual function, to be considered for vision assessment.

Overall, we found a high level of agreement across runs of the same test (averaging between participants: μ=0.01, σ=0.05), as well as consistency across the contrast-BP estimates from tests #1 to #6, resulting for each participant (the mean correlation index, between the six indexes found respectively by correlating the individual estimates on each test with the mean-BP on the remaining tests, was strong, R=0.84, and significant, $p<0.01$).

In summary, when looking at the average performance on each test, across the four groups (Figure 44 and Table 13), we notice that: test#1 shows a strong reliability (reduced only for C+A), as much as tests #2 and #3, but, in contrast, a more limited

validity in detecting individual differences. Test #2 efficiently differentiates groups characterised by un-equal SEDs (it scored the best F static) and maintains the inter-observer differences in performance on repeated measurements (best MAP), but it lacks in estimates' reliability (ICC and CoR are poorer than e.g. for both #1 and 3). Test #3 exhibits a good overall compromise between reliability of measurements and validity in estimating the individual SED (CoR and ICC best results; despite the poorer results for C-A compared to those for other groups), also when test-retest reliability is calculated considering inter-subject variability to weight the individual estimates (best FRP). In general, tests #4, #5, #6 and #7, show poorer reliability and validity compared to the other tests.

Before we discuss the results in more detail, *why did we focus on binocular SED?* For instance, one might want to evaluate sighting-dominance, to assess which of the two eyes dominate fixation. However, as introduced in Chapter 5, different tests for sight-dominance could produce opposite results. Moreover, sighting- and sensory-eye often do not correspond (Jonathan S. Pointer, 2012), and only SED quantifies binocularity on a non-dichotomic scale. In addition, patients with strabismic amblyopia exhibit a clear sight-dominance for the non-deviated eye, making further assessment redundant.

Here we wanted to extend the investigation of the series of tests introduced in Chapter 5 to amblyopic patients, and to further validate the paradigm that test #3 is based on (that we used to monitor BBV treatment; Chapter 3). Based on the definition of the amblyopic syndrome, we expected those participants to have different monocular acuities and, possibly, different contrast sensitivity and gaze-stability in each eye (McKee et al., 2003). Therefore, methods that measure SED based on a comparison of monocular-performance (e.g. C. Suttle et al., 2009) would not have been suitable for this study as the performance would be biased by the level of baseline vision. The same would be true of sight-dominance testing, which would also be less informative than SED in that it does not quantify eye-dominance. We could measure monocular-SED or sight-dominance *in addition to* SED, but due to time limitations we decided to focus only on SED and stereoacuity (the current gold standard measure of binocularity in the clinics).

Stereoacuity. When focusing just on the results presented in this chapter (A+A, C-A, C+A; without considering A-A), test #8 produced inconclusive results for most of the participants (associated results were excluded from the analysis). Stereoacuity assessment is notoriously unreliable, especially in children and especially when stereovision is poor. For instance, 25% of children who fail the Randot test will pass

when retested and, generally, an improvement of ~1 test level is expected on a second testing (Adler, Scally, & Barrett, 2012). To try and improve test reliability we administered a minimum of 5 training trials (more when necessary) to each participant. However, we obtained a complete set of two stereoacuity for only nine children (four with amblyopia) of the total 30 participants. The remaining children and all adults tried to complete at least one run, but none progressed beyond around half of the total planned trials.
Might the duration of the target, in our case limited to 4s, have contributed to poor performance? It has been shown that stereopsis can reliably be estimated provided exposure duration exceeds 200-400ms (Westheimer, 2013). Our exposure duration was selected to respect this limit, giving the participant enough time to comfortably view the target but at the same time limiting the total testing time.

According to the individual results from the last visit recorded in clinical notes, none of our adult amblyopic participants had measurable stereopsis. One might argue that they therefore should not have been recruited to investigate a stereoacuity test. However, there is evidence (for a minireview see Heron & Lages, 2012) for widespread insufficient screening of binocular vision (often just a single measurement of stereoacuity), and that unreliable results on stereopsis - e.g. form-in-depth discrimination (required for instance to solve Frisby or Titmus tests) may result from e.g. patient's head movements during testing. We screened potential participants based on their ocular motility and cover test results, but did not exclude anyone based on a lack of stereoacuity. Only two adults reported inconsistent results from a repeated cover-test (and cases of alternating strabismus were excluded, as stated in the protocol). Therefore, we consider it unlikely that those whose stereoacuity was not reported (7/10; for the other three, the notes cite 'inconclusive' results) could not achieve fusion.

Order-effects? Tests were conducted in a fixed order from #1 to #8 with the goal of making the test-sequence as non-stressful and as engaging as possible. We started with an intuitive test, #1, presenting non-rivalrous letter-stimuli that relied on a letter-identification task, which should result familiar as it reminds of any typical clinical test. In this way, we also hoped to increase a focused participation, by eliciting that 'sense of seriousness' typically attributed to any clinical assessment. We built on tests' difficulty by proceeding to test #2 (form-rivalrous letters) and to test #3 (polarity-rivalrous letters). We moved on to more unfamiliar stimuli incorporating gratings: test #4 with non-rivalrous gratings followed by #5 with rivalrous-gratings. Test #6 move on to dynamic-rivalrous stimuli, which appeared "patchy" to many participant, making the judgement more challenging. Test #7 replicated #6 but had participants use the keyboard themselves to

record their responses (which added a level of complexity). We decided to leave the unique test #8 to the end of each sequence to (a) increase the variety of tests (and maintain attention) and to (b) avoid 'needless' discomfort (possibly caused by the difficulty in completing this test). In fact, throughout the testing, by selecting this order of the tests, we sought to maximise participants' engagement and minimise the impact of fatigue.

Tests # 4 to #7 resulted in an overall poorer performance than the previous tests. Could this poor performance result from the order with which tests were administered? We consider it unlikely that a given test would interfere with the test that followed it. With only 20 trials per test, tests with static stimuli (#1-5) took no longer than 1.5 min to complete. Additionally, around ~1-2 min elapsed between two consecutive tests (while the experimenter set up the next test, provided instructions and answered questions). Thus, these aspects combined should have eliminated the possibility of one test influencing results from a subsequent test.

Practical aspects of testing. We note that not all tests were equally easy to explain and children, in particular, struggled in understanding phase-discrimination tasks (#4-5). Further, poor acuity (e.g. in the C+A group) compounded the difficulty some participants had in performing these tasks, which likely contributed to their generally poorer reliability. We speculate that these tests may require more familiarity with the task and a more sustained attention that tests #1-3 to provide results equally reliable and valid as from those tests.

Which stimulus-feature should be manipulated? Stimulus features, such as size and contrast, can influence performance when testing binocular vision in terms of either binocular-rivalry (O'Shea, Sims, & Govan, 1997), stereopsis (Heron & Lages, 2012) or binocular-combination (Bosten et al., 2015). All these processes likely contribute to sensory-eye-dominance (SED). We used tests involving binocular-*combination* of stimuli that were either overlapping in space (#2, #3, #4, #5, #6) or did not overlap (#1) and we had some tests requiring *suppression* (#3, #5), one test measuring *rivalry* (#7) and another measuring *stereopsis* (#8). Comparing individual scores across tests, the results we obtained were consistent within participants, which supports the hypothesis that these phenomena are supported by a common neural mechanism. However, results differed between participants (comparing individual scores on the same test), indicating a range of individual differences in performance, more or less widely spread from a task to another. Moreover, the magnitude of SED was not directly associated with the

individual level of baseline acuity and/or inter-ocular difference. Possibly, different binocular-vision processes are sensitive to different features of the stimuli.

We have highlighted how binocular combination is affected by contrast modulation (tests #1-6) & luminance (test #3), using static (tests #1-5) and dynamic stimuli (test #6). We modulated the contrast of dynamic stimuli to test binocular-rivalry (#7) and, when a measure could be obtained, we quantified stereopsis by varying the horizontal disparity of high contrast stereograms (test #8). The specific identity of both elements of each pair of dichoptic stimuli was either a letter (tests #1, #2, #3, #8), a horizontal grating (shifted in phase; tests #4, #5) or a pattern of 'noisy-dots' (tests # 6, #7). Associated tests tapped into form, position or motion discrimination. Considering both individual differences in binocular vision (Bosten et al., 2015), that have specifically been demonstrated for binocular rivalry (Dieter et al., 2016), and the variety of functional visual deficits associated with amblyopia (E. E. Birch, 2013; J. M. Holmes & Clarke, 2006; McKee et al., 2003), a series of tests may be the best way to assess binocular vision in amblyopia. This would facilitate vision assessment, by overcoming i) response-uncertainty, related for example to familiarity with the task, attention etc, and ii) uncertainty about eye-dominance, e.g. as influenced by the magnitude of eye-dominance (Yang, Blake, & McDonald, 2010). On the latter point, we note that eye-dominance might not be fixed when presenting certain rivalrous stimuli. In fact, cases of 'Mixed-dominance' have been shown to derive from dichoptic-rivalrous-stimuli being more or less reciprocally similar in colour and size (O'Shea, Parker et al., 2009). Also, in tasks involving visual competition different results in repeated measurements are likely to be obtained (Blake & Logothetis, 2002).

SED and amblyopia There is increasing interest in the feasibility of treating amblyopia in older children and adults, either with occlusion therapy (Erdem, Cinar et al., 2011; Kishimoto, Fujii et al., 2014) or other monocular (Evans, Yu et al., 2011) and/or binocular regimens, such as dichoptic training and video-game play (for a review see Tsirlin et al., 2015). Based on the idea that inter-ocular suppression is a cause, not a symptom of amblyopia it has been proposed that binocular treatments achieve their results by reducing inter-ocular suppression. However, a few studies now suggest that therapeutic benefit cannot be underpinned solely by a reduction in inter-ocular suppression (Bossi et al., 2017; Vedamurthy, Nahum, Huang, et al., 2015). There are a number of current hypothesis about the mechanism supporting this improvement in vision but importantly there are also a large number of different measures of SED in use across different labs.

The assumption that all tests of SED tap the same mechanism may not be true and this may hamper comparative evaluation of the outcomes of different binocular therapies.

6.4.1 Conclusion

Measures of SED have a role to play in quantifying binocular capacity in the absence of measurable stereoacuity, as frequently occurs in in amblyopia. The current study compares on the one hand individuals with and without abnormal binocular vision, and on the other compares adults and children. Participants in groups A+A, C+A and C-A exhibited a wide range of BPs, generally corresponding to higher SED-magnitude than those measured for A-A. This manifest as an advantage for the less sensitive tests that previously showed strong test-retest reliability. Test #1 was able to differentiate BPs across observers more than it did in A-A; test #2 appeared very useful in adult patients, when maybe the rivalrous letters facilitated their recognition (although at the expenses of showing stronger eye-dominance); test #3 again had high sensitivity, manifest as either consistency in repeated measurements, or as precision in highlighting individual differences in performance. Phase tests (#4,5) revealed procedural difficulties. Tests #6 and #7 require further investigation: so far they appear useful as an *additional* indication of SED for dynamic stimuli. An accurate and reliable evaluation of binocular vision (not just in amblyopia), should be quick - as other methods, e.g. (Yang et al., 2010) - but also capture the range of visual abilities/processes involved in binocularity, i.e. not be limited to either binocular combination or rivalry.

In presence of amblyopia, having a standard set of tests to measure binocularity would be useful in clarifying how and in what sense inter-ocular suppression, as quantified by the magnitude of imbalance in SED, affects therapeutic outcomes for amblyopia and the mechanism that lead to this disorder.

7 General Discussion

Overview

We started by evaluating the efficacy and safety of a novel binocular treatment for childhood amblyopia. We evaluated therapeutic outcome (primarily, acuity) and determined which factors, e.g. age of participants, severity/type of amblyopia, influenced the results. In addition to clinical outcome, we measured performance, pre- and post-treatment, on three psychophysical measures (acuity, contrast and crowding) known to be compromised in amblyopia. Our findings did not indicate that treatment reduced eye-dominance, although we did find a general improvement in acuity. Can the success of a therapy for amblyopia be solely explained by its anti-suppression effect? To help answer this, we also investigated the types of suppression-measures available, and how they relate to our own measure of suppression. In this way, we sought to determine the extent to which our own findings depended on the type of suppression measure used.

Overall, this thesis sought to examine the following questions:

- Are binocular treatments for amblyopia a valid alternative to traditional occlusion?
- If so, how do they work?
- Can we develop better methods of measuring sensory eye dominance for evaluating binocular visual function (eventually, following treatment)?

This project has focused on *unilateral* amblyopia, diagnosed in the presence of monocular strabismus and/or anisometropia (see Chapter 2) resulting in ≥0.2 logMAR inter-ocular difference in acuity (with BCVA in AE≥0.3 logMAR). A universal diagnostic standard does not exist for amblyopia and different studies have used different criteria to define amblyopia, complicating the comparison of results. For instance, it is difficult to define *positive* diagnostic criteria (i.e. not defined by exclusion) for amblyopia. In fact, the deficit manifests through a variety of symptoms (K. R. Kelly et al., 2015; McKee et al., 2003), including reduced form recognition (acuity; the most clinically relevant symptom) as well as a reduced contrast sensitivity, abnormal vulnerability to foveal crowding, fixation instability and compromised binocular-vision.

Treatment of amblyopia varies widely (David K. Wallace et al., 2018). Monocular occlusion treatments have been the gold standard since their introduction over a thousand years ago (for more on the history of amblyopia treatment, see Loudon & Simonsz, 2005). Despite their well-known limitations – the emotional stress induced by wearing a patch in public, poor compliance, limited impact on binocular vision - and alternatives being increasingly widely available, patching and tropical atropine remain the default treatments offered to patients (V. Tailor et al., 2016). In addition to the drawbacks of occlusion therapies, increasing evidence about the relationship between amblyopia and binocularity (for an overview: E. E. Birch, 2013), is inspiring researchers to explore the development of new treatments. These treatments use binocular treatment methods as *independent* tools or additional *aids* to traditional-monocular occlusion (see section 2.8). Different techniques have been proposed some of which have been thoroughly explored in randomized clinical trials (for an overview, see Table 6). At present, there is no clear agreement as to the superiority of one treatment method over another.

We have presented a new binocular therapy for amblyopia (Chapter 3). Although the results we report for children are promising, we cannot conclude whether this method has specific therapeutic advantages over other treatment methods, and in particular, over other binocular therapies. A Phase II UK-based clinical trial is now underway, to compare our treatment to traditional patching. We have however confirmed the association between amblyopia and *crowding* (results are discussed in Chapter 4) and have explored a measure of visual function potentially useful for the clinical diagnosis of amblyopia: *sensory eye dominance* (SED). By definition, amblyopia is characterised by some degree of inter-ocular imbalance in visual function: at least a 2 logMAR line-difference in best-corrected acuity between the two eyes (J. M. Holmes & Clarke, 2006). Therefore, it would be reasonable to expect higher levels of imbalance, i.e. a greater SED, in amblyopic patients compared to controls (results are respectively discussed in Chapters 5 and 6). Whether measures of vulnerability to foveal crowding *and* SED might facilitate early diagnosis of amblyopia (in children) remains to be seen.

Visual crowding is already considered a critical component of visual assessment, especially but not exclusively for patients with amblyopia. Indeed, the testing of crowded-acuity is generally recommended from as soon as the child is able to perform the task - children generally consider it more difficult than a single letter/symbol recognition (J. M. Holmes et al., 2001). This strategy aims to reduce over-estimation of acuity (Anstice & Thompson, 2014) to avoid prescribing a correction that might be insufficient in a natural

(crowded) environment. We suggest that SED might be an informative measure to add to a vision assessment, as a useful addition to stereoacuity as a means of evaluating binocularity. For instance, unlike stereoacuity, we found that SED was measurable on patients regardless of the severity of their amblyopia.

7.1 Summary of findings

7.1.1 BBV treatment (clinical and psychophysical results)

In Chapter 3 we presented a novel home-based treatment for amblyopia: Balanced Binocular Viewing (BBV). The treatment required that participants wear shutter glasses, to allow for dichoptic stimulus-presentation, and watch modified videos on a 3D-capable computer monitor. The AE then received an un-modified image, while the FE received a filtered-version of the same image, blurred sufficiently to match the individual crowded acuity of the FE to the (crowded) acuity measured in the AE. Content presented in this way should be equally visible across the eyes. Every day, the patient was given a choice of viewing from a series of cartoons, movies and educational content, each ~1hour long. We believe that this variety promotes adherence to the prescribed daily dose (1 hour-watching). As a safety measure, i.e. to monitor the effect of BBV on inter-ocular suppression, and to monitor compliance, a simple game (the “ghost”-test) was played throughout movie-viewing. For this study, treatment efficacy was assessed based on the results obtained during regular orthoptic assessments, at the hospital, after 4, 8 (mandatory) and 16, 24 (optional) weeks from baseline, depending on the individual response to BBV. The treatment was interrupted during or after the 8-week visit if a minimum improvement of 0.1 logMAR (widely accepted limit for acuity test-retest variability; J. M. Holmes et al., 2001) was *not* obtained from the previous visit.

Twenty-two children, with anisometropic (n=7; Group 1) and strabismic or combined-mechanism amblyopia (n=6 and 11; both in Group 2) participated in this study. On average, children watched 54 ±14.5 minutes of videos per day, on more than two-thirds of the days on which the equipment was available to them. Overall, the acuity in the amblyopic eye improved by an average 0.27 logMAR (σ=0.22). In general, this gain was not correlated to a reduction in inter-ocular suppression. Neither the individual gain in acuity nor the change in IOS was significantly influenced by the different durations of treatment, i.e. 8, 16 or 24 weeks. The different duration was based both on the protocol, i.e. max 8 or 24 weeks respectively for children in Group 1 or 2, and on the individual response to treatment, periodically controlled by an orthoptist, during the follow-up visits.

Age, interocular-acuity-difference at baseline and type of amblyopia also did not significantly influence acuity-outcomes.

To further investigate the pattern of visual loss pre- versus post-treatment, we collected a series of monocular psychophysical measures of acuity, contrast sensitivity and vulnerability to crowding. The results (Chapter 4), confirmed that BBV improved acuity, although the mean gain was 0.05 logMAR lower, on average, than gains calculated from acuity measures made during orthoptic assessment (i.e. with crowded charts). Results for acuity compared to those for crowding revealed a trend (although not statistically significant) towards a reduction in the susceptibility to crowding being more marked when the gain in acuity was greater. We also collected psychophysical measures for stereoacuity, but results were inconclusive. This was consistent with the inconclusive clinical results, collected using Frisby test during follow-up visits; only 8 children who completed BBV demonstrated measurable stereopsis (one of them progressed from 'unmeasurable' stereopsis at baseline). However, 7/8 children showed a significant improvement in stereoacuity. Thus, when stereoacuity was measurable, our data revealed a positive effect of BBV on binocular vision. This leaves open the possibility that a change in the other children might have occurred too but was *not* detected during their vision-assessment. That the success of treatments for amblyopia in improving binocular vision is usually evaluated based purely on stereoacuity-results further motivated us to explore different methods to assess binocularity. First, we had to validate our 'ghosts-test', the self-administered test of sensory eye dominance built into the BBV therapy. Results from that test remained, on average, unchanged over the course of treatment, but would this result be confirmed using other (in theory) equivalent tests? Second, we also wanted to compare different methods that could quantify binocularity in any patient, especially when stereoacuity is not measurable. That study is described in Chapters 5 and 6.

Strengths and weaknesses (see also section 3.4)

Lack of a control group. Recall that we classified the BBV study as exploratory (see section 3.1) and did not have a control group (i.e. one that received a sham treatment). This design reflected our limited resources, in term of time (to complete OT and the assigned treatment) and of funding (to support additional clinical staff, e.g. for recruitment, and to provide more PC-systems for the therapy). Nonetheless, we tried to isolate the effects of BBV, 'controlling' for the individual history of treatment for amblyopia. Note that our intention was primarily to explore the safety and efficacy of BBV.

A recent meta-analysis of the effects of PL, video gaming and dichoptic training on adult amblyopes, found that any treatment methodology was similarly effective in improving acuity in the AE, i.e. no one task and/or viewing condition was better than others (Tsirlin et al., 2015). If so, a useful control for our study would have been a sham-treatment group, to be treated together with the real-treatment group, to test the efficacy of the specific treatment itself (knowing that any treatment might be similarly effective). However, playing at modified dichoptic videogames has been shown either to be specifically more effective (E. E. Birch et al., 2015; S. L. Li et al., 2014) or to provide a non-superior treatment of amblyopia (T. Y. Gao et al., 2018; Nicola Herbison et al., 2016) than playing at placebo videogames. However, Tsirlin et al. also reported that a higher initial severity of amblyopia was correlated to greater improvements after treatment, in almost any study reviewed. The incidence of the severity of amblyopia is well-known to impact also the effects of occlusion therapies (Pediatric Disease Investigator Group., 2003; Stewart et al., 2005). Therefore, another important control regards the history of prior treatment, before enrolment into the investigated study. For example, if a minimum OT period is not controlled for, there might be more space for improvement, before vision stabilises. The lack of control for OT might for example explain why anisometropic, more than strabismic/combined amblyopes, would improve because of a period of full-wearing the updated refraction. So, in contrast to what has been previously claimed (Vedamurthy, Nahum, Huang, et al., 2015) the different response from participants with a different type of amblyopia (i.e. showing only, also or no anisometropia) could be elicited by the OT period independently of the parallel treatment modality in itself (we discuss this in the next paragraph). Given we controlled for OT, should any other prior-treatment be controlled for? As stated above, Gao et al. found that older children (>7yrs) and adults (where ¾ had had previous patching or atropine treatment), assigned to either a specific treatment modality or to a placebo equivalent modality, reacted to a period of 'treatment' in a similar way (T. Y. Gao et al., 2018). While another recent trial found that children (4-9yrs), irrespective of some being pre-treated (wearing a patch) and some untreated, received equal and greater benefit from binocular game-play than from standard occlusion (Krista R. Kelly, Jost et al., 2018). This study follows from E. E. Birch et al. (2015), who demonstrated a superior effect of this binocular treatment compared to its sham-version. The authors of the last 3 mentioned studies have all controlled for prior full-OT period but only Kelly et al. considered if prior occlusion had an impact. However, their study mixed monocular and binocular treatment modalities, before and during the study. Thus, a possible 'ceiling' for the cumulative effect of different therapies has not been ruled out. To compare the specific effect of a binocular therapy

to those specific from traditional monocular therapy, first we have explored the isolated effects of BBV treatment (binocular; which impact has been explored in children who had previously received OT but no other treatments for amblyopia). This exploratory study has then piloted the aforementioned controlled clinical-trial. So now, we are aiming at comparing BBV to patching (monocular therapy; commenced after a period of OT, in children of a similar age as those participating to BBV and, as well, having no history of other previous treatments for amblyopia).

The impact of OT period. We only included children with untreated amblyopia, i.e. having received no treatment other than a minimum of 16 weeks optical treatment (OT; also called 'refractive adaptation'). OT consisted of full-time wear of prescribed optical correction until best corrected-acuity "plateaued". The minimum period for OT was set to 16 weeks based on prior evidence suggesting that this is sufficient to achieve best visual acuity, independent of age and type of amblyopia (precisely, the BCVA has been found to be achieved after 14.7 weeks of monitored OT; Stewart, Moseley, Fielder, et al., 2004). Note that in children 3 to ~7 years, a period of 18 weeks of OT, alone, has been found to improve vision (≥2 logMAR lines improvement in the AE acuity) in ~2/3 cases of amblyopia (Cotter et al., 2006; Cotter et al., 2012) and to resolve amblyopia (i.e. AE acuity improves to within 1 line of the FE acuity) in ~20-30% cases of anisometropic (Cotter et al., 2006) or strabismic/combined amblyopia (Cotter et al., 2012). Following a shorter average period of OT, other studies have reported lower proportions of resolved amblyopia, in children aged 3-7yrs (7%, after 14±3w. OT; Stewart et al., 2007) and of improved amblyopia, in children and adults (6-9% at near and distance acuity, after 4-16w. OT; T.Y. Gao et al., 2018). The average 'time to best acuity', yielding a mean improvement of ~2.5 logMAR lines in AE acuity, has been set around 14-15 weeks (Stewart, Moseley, Fielder, et al., 2004), while ~8 weeks has been shown to be enough to stabilise vision in 90% cases and to gain at least 1 logMAR line (T.Y. Gao et al., 2018). Surprisingly, the positive effect of OT seems to be similar in the presence or absence of strabismus and is not influenced by the relative different age, either between children (Moseley et al., 2009), or between children and adults (T.Y. Gao et al., 2018). Monitoring the OT period significantly improves its impact (Stewart et al., 2007), which could explain the different proportion of recovered deficit, reported above. In our study, we relied on subjective report that children wore their glasses full-time: for 36/40 children approached (discussed in section 3.2.1.2), a mean OT period of 26±10 weeks lead to a mean substantial improvement in the AE acuity of almost 2 logMAR lines (μ=0.17 logMAR). Despite the high variability (mean σ=0.28 logMAR), we found

that amblyopia resolved in 5% and improved in 57% of children (with 46% showing ≥2logMAR lines improvement in the AE acuity).

We reaffirm that it is essential to control for the effect of OT when evaluating the efficacy of any treatment for amblyopia. Recall that the deficit is defined as 'a *persistent* difference in the *best-corrected* acuity [in absence of other ocular pathologies]'. With that in mind, that we only included untreated amblyopes allows us to isolate the effects induced by BBV, and limited potential 'cumulative'-therapeutic effects that might have derived from BBV *and/or* from different previous therapies. In this way, we can be sure that outcomes derived from our therapy (after acuity had stabilized following OT).

The disadvantage of selecting children who had no previous treatment is that we cannot rule out the idea that receiving *any* treatment might trigger a change, interrupting so-called 'pathological equilibrium'. We note that among the 13/22 children who received patching after being released from BBV, the mean gain in the AE acuity at treatment completion was only marginally different than that measured after ~1 additional year (paired t-test $_{(df12)}$: p=0.074). This suggests that the introduction of another therapy had only a minor additional advantage and that it was probably BBV that triggered the initial change. However, our design means that we cannot exclude that e.g. patching followed by BBV would have produced similar results.

We included children from aged 3 to 11 years. This means that we explored the effects of BBV during and just beyond the most 'sensitive period' to treatment for amblyopia, widely considered to last until a child is around 7 years old (J. M. Holmes et al., 2011). The fact that we elicited high levels of compliance both in fickle participants (young children) and in those who might be reluctant to wear a patch (~7-11 yrs old children), and that the level of compliance did not depend on the age of children, suggests that BBV would be amenable to children of all ages. However, the efficacy of BBV may be reduced in older children and adults, who fall outside of the 'sensitive period'.

The primary outcome. Any treatment for amblyopia is typically evaluated on the basis of the final best corrected monocular visual acuity which remains the principal criterion for the diagnosis of amblyopia and the main objective for any treatment plan (David K. Wallace et al., 2018). We relied on measures obtained on crowded logMAR charts (Thompson v2000 software; Thompson Software solutions, Hertfordshire, UK), by a clinician. We found that gain in acuity was not related to the type of amblyopia and, at the same time, was not linked to a significant change in the level of eye-dominance

(quantified using the ‘ghost’-test). The neural mechanism thought to be responsible for sensory eye-dominance is usually termed ‘inter-ocular suppression’ (IOS) or simply ‘suppression’, and indicates the adaptive mechanism whose purpose is to limit the burden of amblyopia, e.g. by limiting fusion of binocularly un-balanced inputs to avoid diplopia (Spiegel et al., 2016).

Our findings challenge the belief that suppression is necessarily the mechanism targeted by a successful therapy for amblyopia (R. F. Hess, Thompson, et al., 2014; Krista R. Kelly et al., 2018; J. Li et al., 2011; Narasimhan et al., 2012). In fact, if a therapy acted directly on reducing suppression (reducing the inhibitory role of the FE input while stimulating the excitatory role of the AE), then binocular training should have had no effect on AE-vision unless the dominance of the FE reduced during the treatment period, by at least an amount that would be enough to allow the input from the AE to be used. In our study, even though IOS remained largely unchanged, AE monocular acuity improved similarly across all children, independently of the type of amblyopia. This suggests that the level of IOS does not directly affect acuity and, conversely, that deficits in acuity might be not particularly indicative of a deficit in binocular vision. Instead, there appears to be a complex relationship between these measures that are likely influenced by strabismus (Kehrein et al., 2016). Thus, we argue that IOS cannot be the sole mechanism responsible for amblyopia and we discuss alternative hypotheses in section 3.4.

Compliance. The BBV treatment was engaging to children and popular with parents. During the post-treatment home-appointment (for us to recover the therapy-PC, screen etc.) we provided parents with a questionnaire to be sent back to the experimenter. Based on the opinions of 7 parents, we know that the system did not have a negative impact on family life, and instead kept the child (and siblings sometimes) occupied and watching the TV-screen for at least half an hour, which the child enjoyed. Although most children viewed content more than once, they remained attentive to the content as we did not report, on average, a high rate of ‘missed’ or ‘not reliable’ answers to the ghost-task (which required a response every minute). Moreover, we did not find a significant correlation between compliance and improvement in acuity, which suggests that the advantage of exceeding a certain daily dose brought no further benefits in terms of acuity-gains. Exactly what the optimum daily dose is, remains to be investigated. Many factors should be taken into account, such as age, severity and type of amblyopia, in a further study that would also include a control group. Note that a possible explanation for a non-significant correlation between higher compliance and greater acuity-gain

might be that all children demonstrated a uniformly high level of compliance, with a low variability across participants that led to having an insufficient variation to power a correlation. For a future investigation, e.g. increasing the sample size will reduce the probability to incur in the same situation.

We note that in most cases, what we required our patient to do (i.e. spend one hour a day watching a movie or cartoons, while playing the 'ghost'-game), would not have constituted a major departure from the child's regular routine; and parents are not generally concerned with 1hr viewing-time (Christakis, Ebel et al., 2004), i.e. the daily dose we prescribed. However, we specifically required this routine be adhered to on a daily basis, in contrast to a general 'tendency' to watch TV daily. Children complied well, watching at least 20 minutes a day on 68% of days on treatment; but with a period of holidays in between, meaning that sometimes they missed e.g. 6 days but were then following the prescription for the following 12. As we did not have a sham-treatment control group (e.g. daily watching, wearing the goggles but watching *un*-modified movies) we cannot conclude whether it is the *perseverance* in watching movies *(every single day)* or the specific modality of movie-watching (binocularly balanced-viewing) to positively affect acuity. i.e. to have the therapeutic effect. .Previous studies have either *discounted* the idea that a daily sham-activity (placebo game-play) had the same effect as the daily actual-treatment-activity (modified dichoptic game-play; E. E. Birch et al., 2015) or *confirmed* that the two activities made no difference, provided a similarly good daily compliance (T. Y. Gao et al., 2018). This leaves open the possibility that simply engaging in a controlled daily visual activity might elicit a vision improvement.

Adverse events. Treatments for amblyopia rarely elicit serious adverse event. However, the lack of measurable stereopsis frequently observed at the start of treatment combined with improvements in acuity that arise from treatment, means that any therapy needs to take seriously the risk of inducing diplopia. This is especially the case in the presence of strabismus, and even more so if the angle of deviation significantly changes throughout the treatment. We note however that intractable diplopia as a result of patching is rarely reported (e.g. only 24 cases in 5 years in the UK; Newsham & O'Connor, 2016). Further we note that strabismic participants showed, on average, only minor and not-significant changes in their angle of deviation (Δ pre- versus post-BBV; paired t-test$_{df13}$: p=0.25), in accordance with previous evidence for acuity improving despite the unchanged inter-ocular (mis-) alignment (Cotter et al., 2012). Predictably then, we did not observe any cases of diplopia arising from BBV treatment. This is based on patients'/carers'

subjective reports, or on Sbisa measurements (when obtainable and/or reported in the clinical notes).

‘Reverse amblyopia’ (also called ‘occlusion-amblyopia’) might occur after extensive treatment for amblyopia, as degraded vision in the non-treated eye (FE), whose acuity would reduce beyond baseline. However, so far, cases of reverse amblyopia have been rarely reported and only following monocular (not binocular) treatments, in particular atropine (5%; Hainline, Sprunger et al., 2009), that lasts uninterrupted for days, depending on the dose. As well, patching-regimen must be controlled to avoid similar adverse effects (J. Holmes et al., 2004; Longmuir, Pfeifer et al., 2012). So far, binocular treatments have not demonstrated similar adverse effect. In support to this, also after ~50 weeks from BBV treatment cessation and the cumulative effect with further monocular treatment, we report no cases of reverse amblyopia.

Although we required sustained attention to a daily activity, most children reported enjoying the therapy as demonstrated by their good compliance. None reported eyestrain. We note that some children did find the pre- and post-therapy psychophysics quite taxing, as demonstrated by the fact that many children struggled (or were not willing at all) to pay the required attention for the duration of all the tests. We did not report the same inconstant behaviour during the daily sessions of BBV treatment.

7.1.2 SED in normal and amblyopic viewers

When the two eyes have similar refractive-power and converge on the same target with enough precision, sensory-fusion and motor- fusion normally occur. These processes are limited: the former by individual spatial resolution (C. Schor et al., 1984), the latter by objects’ spatial configuration in the binocular visual field (Vojnikovic & Tamajo, 2013). These phenomena support normal binocular vision: observers perceive a coherent cyclopean percept and the relative depth between objects in the scene (Julesz, 1971). Thus, good stereopsis is only achieved in the presence of fully functional fusion. Consequently, stereoacuity should in some sense be considered as the finest level of binocularity but not the only one, so not the only meter to assess binocularity (e.g. fusion can be present but be ‘poor’, resulting in a sensitivity to retinal-disparity too low to support stereopsis).

An additional measure of binocular visual function consists of assessing the relative contribution of each eye to the binocular percept, i.e. to quantify *sensory-eye-dominance*

(Ooi & He, 2001). Under dichoptic viewing, the relative strength (e.g. contrast) of concomitant images supporting optimum performance is termed the interocular *'balance point'* (BP; e.g. contrast-BP). The BP quantifies the extent to which an observer relies on the two components of a dichoptic stimulus-pair. Typical stimulus-pairs used to explore this process are dynamic random-dots patterns, one of which contains a *coherent-motion* component that must be detected while the contrast of the other *random-motion* component is increased (Black et al., 2011; J. Li et al., 2010; J. Zhou, Jia, et al., 2013). A second category of tests uses rivalrous stimuli to quantify SED. When different monocular images fall on corresponding retinal locations of the two eyes this can lead either to an experience of *binocular rivalry* (an alternation of percept between the stimuli presented to the two eyes; Wheatstone, 1838), or to binocular diplopia (sense of double vision; Rucker & Tomsak, 2005). Considering amblyopia as a deficit of binocular vision, we adapted and developed a series of tests that could serve as a reliable and quick way of quantifying binocularity in the clinic (see sections 5.1 and 6.1).

Summary of the main findings.

We compared the performance of 40 adults (10 with amblyopia) and 20 children (10 with amblyopia) on eight tests that use dichoptic stimuli to quantify binocular visual function, both in term of SED (tests #1-7) and stereoacuity (#8) using stereo-defined Sloan letters (variant of Frisby et al., 1996). All SED tests involved estimation of a balance-point (BP). Specifically, pairs of *contrast*-modulated stimuli were used in tests #1-6, comprising Sloan letters (adapted from Bossi et al., 2017 -test #3; Kwon et al., 2015 -tests #1-2), reciprocally-shifted gratings (from Ding & Sperling, 2006 -tests #4-5) or up/down drifting noise-patterns (loosely based on the motion task by Black et al., 2011 -test #6). Test #7 (inspired by Dieter et al., 2016) measured rivalry, as the proportion of dominant-eye instances (presenting the same pairs of stimuli used in #6, but at a fixed contrast). We assessed SED scores based on 1) *Reliability* (consistency between two measurements, made over two runs on the same participant) and 2) *Validity* (the ability of the test to capture individual differences). Overall, tests #4 to #7, showed poorer reliability and validity compared to the other tests, and were less intuitive to participants. In contrast, tests #1, #2 and especially #3, showed stronger validity and reliability compared to the other tests. Interestingly, we notice that test #3 had the advantage that patients could indicate the position of the 'whiter' letter (e.g. reporting 'the one on the right'; instead of having to recognise the specific letter, as required by the tasks in #1, although letters were non-overlapping as in #3 and in #2).

SED as part of vision assessment.

Our pilot study (Chapter 5) investigated this series of tests on 30 adults with ostensibly normal vision (no amblyopia; Group A-A). Interestingly, we found no significant correlation between stereoacuity and SED. Increased binocular balance (i.e. BPs around 0.5) was not associated with better stereoacuity. This suggests that these phenomena might be supported by a different mechanism and possibly might have a different course of recovery during treatment for a disorder of binocular vision. To further examine this idea, (in Chapter 6) we extended the investigation to 10 adults with amblyopia (Group A+A) and 20 children, 10 with- and 10 without- amblyopia (respectively, Groups C+A and C-A). Unfortunately, only 9/20 children (four from C+A) and none of our amblyopic adults (A+A) completed two runs of our stereoacuity test (#8). For these nine children, results suggest that test #8 has the potential to be a valid index of stereoacuity. The scarcity of stereoacuity data (we obtained results for only 9/30 participants across Gr. A+A, C+A and C-A) did not allow a complete comparison to our SED-tests specifically in term of validity and reliability, nor across tests, nor between observers' groups. However, these results provide an example of how relying wholly on stereoacuity when evaluating binocularity can be limiting.

Looking at the SED tests across the 4 groups of observers, we found that the presence of amblyopia significantly influenced both the amount and the distribution of mean individual SEDs. Our tests were sensitive to the presence of amblyopia, with the mean inter-ocular imbalance in contrast sensitivity being greater in amblyopes compared to non-amblyopic participants by about a factor of 3. Across A+A and C+A, our SED indexes have matched the side of amblyopia, i.e. the mean BP varied from 1 to 0 in accordance with the right-eye (0.5-1.0) or the left-eye (0.0-0.5) having a better visual acuity. For the contrast-BP tests (#1-6), we found (averaging between participants) a high level of agreement across runs of the same test. We also report consistency across the six contrast-BP estimates within participants (here, all 60 participants, from all four Groups). In addition, by averaging across all the participants exhibiting an inter-ocular difference exceeding the acuity test-retest variability (i.e. ≥0.1; J. M. Holmes et al., 2001), we could show that the *amount* of SED was *not* significantly correlated either with the level of monocular-acuity (severity of amblyopia) or with the interocular-difference-in-acuity. Thus, SED is not necessarily linked to the resolving power of each eye. In other words, SED represents a separate aspect of vision assessment, independent of acuity. Thus, even if adults showed poorer AE acuity compared to children (on average, they read half-line less than children, on log charts), the difference in age did not significantly influence SED. That there are individual differences and, at the same time, some

aspects of vision one would expect to be related might be not (e.g. SED and acuity) suggest there is value in further research to explore how to best redefine and measure the pattern of amblyopia.

Aniseikonia and abnormal-retinal-correspondence

Sensory-eye-dominance is a means of quantifying un-balanced sensitivity (to either contrast or retinal disparity) between the two eyes. Specifically, we quantified SED as a direct measure of the relative strength of each of the two monocular inputs presented on screen, forming a dichoptic-stimulus. However, different people can tolerate small differences in inter-ocular input, due to e.g. optical aberrations, to different extents. It may be difficult to use SED to disentangle un-balanced sensitivity from an individual's tolerance to input-imbalance.

For example, observers tolerate small inter-ocular differences in perceived image-size, that may arise from unequal refractive power between the eyes ('aniseikonia'; tolerance at 5%; Achiron et al., 1997). Consequently, contrast-sensitivity (for example) results from the individual contrast-modulation of the inputs, 'corrected' by the unbalanced magnification in each eye. Observers also tolerate a small offset in retinal image-disparity, arising from a small angle of monocular misalignment ('abnormal-retinal-correspondence'; tolerance is proportional to power of the eye and angle of resolution, σ; it has been set at PD/2cosα; Vojnikovic & Tamajo, 2013). In this case, central stimuli fusion is possible in the presence of micro-strabismus, (<10Δ misalignment) but only within a disparity *gradient*, i.e. depending on the proximity of objects in the visual field (Braddick, 1979). The dependence on spatial objects' configuration could manifest as a higher vulnerability to crowding (preventing stereopsis for even very small disparities if the scene is too cluttered). This tolerance to inter-ocular differences potentially allows stereopsis to occur normally, while binocularity (in term of processing either retinal-image size or retinal-correspondence) is abnormal but may not be noticed by the patient. Specifically, diplopia is not reported and patients are unaware of the fact that rivalry may occur more frequently (where the two percepts interfere with each other but the difference is not perceived) than in observers who did not develop these tolerances. We could think about this form of rivalry as 'suppression associated with no loss of visual information'.

Note that, prior to administration of either BBV therapy or of SED-tests, we did not check for aniseikonia, or for abnormal retinal correspondence. We could, for example, have used a vergence response task using dichoptic nonius alignment, to check for sensory

fusion; and presented it as a self-adjustment task, to check for aniseikonia. In terms of the BBV therapy it is possible that the presence/degree of aniseikonia would have contributed to the responsiveness of patients to treatment. In terms of the SED work, we did try to ensure normal fusion based on screening of clinical notes. With regard to aniseikonia, we recruited the 30 A-A participants through a school of optometry, and participants would i) likely report even minimal discomfort and ii) have received a full eye-examination not long before testing, as part of their course. The other participants were patients with an up to date prescription. The reason we did not check for this is practical: we recruited and tested each participant on a single-session, often in between specialist-visits. Thus, we had to condensate as much as possible the testing session and we dropped this additional control-test based on the belief that clinical notes were complete (and no report of e.g. aniseikonia was noticed). This omission might influence the validity of each tests, but a comparison between tests' reliability would not be affected. Future studies should include extra control-tests for these anomalies.

Sample size.

We were limited (in terms of time and resources) in our ability to increase the sample size, and reduce the risk of type I errors for the law of large numbers. Thus, we selected tests to be as diverse as possible and at the same time, we focused on stimuli involving or not-involving rivalry. Doing so, we tried to limit the probability of our participants to be particularly good in one or another test (maybe, also because of the presence of a binocular anomaly between those described in the previous paragraph). Then, we tried to use statistics that would summarise the main effects we were interested in, while facilitating a simple interpretation of the results, to guide further investigations.

7.2 Implications

7.2.1 Relating current amblyopia research and our findings

Knowledge of the monocular *and* binocular functional deficit in amblyopia have resulted in novel methods to treat amblyopia (Vijay Tailor, Bossi, et al., 2014). Treatments that focus on re-balancing the sensory input across the eyes lead to positive therapeutic effects in adults (Baroncelli, Maffei, & Sale, 2011; Tsirlin et al., 2015; Yao, He et al., 2017), which questions the notion that childhood is the only 'critical period' for treatment of amblyopia (Schmucker, Kleijnen et al., 2010). One advantage of developing treatment methods aimed at balancing visual inputs is their potential for longer-lasting (perhaps

life-long) efficacy, contrary to occlusion treatments, whose popularity declines, while dose-rate increases with increasing age (Jingyun Wang, 2015).

The fact that amblyopia treatment can be effective beyond the critical period for the development of the visual system suggests that there is a degree of *cortical plasticity* in adulthood (Bavelier et al., 2010; J. Zhou, Thompson, et al., 2013). It has been claimed that, in adults with amblyopia, abnormally low cortical activity in some areas can be normalized, for example, using repetitive transcranial magnetic stimulation (Clavagnier, Thompson, & Hess, 2013; B. Thompson, Mansouri et al., 2008) or direct current stimulation (Spiegel, Byblow, et al., 2013).

Practically speaking then, *how could we take advantage of synaptic plasticity to improve vision in patients with amblyopia**?*** We presented our BBV therapy as one of a number of emerging binocular treatments. As reported previously (Kehrein et al., 2016; Vedamurthy, Nahum, Bavelier, et al., 2015), the gain in acuity, measured at the end of the treatment, were not clearly related to a reduction in IOS. Which leads to more questions: *What kind of sensory experience could induce a re-wiring of connections that would solve the binocular-vision deficit in amblyopia? Which binocular-process (or processes; modulating the combined activity of cortical neurons) would be best to stimulate?*

On the one hand, repetitive training on specific *binocular combination* tasks (paradigm 1) has been shown to improve vision. The tasks adopted include either monocular tasks (occluding the non-trained eye), e.g. contrast-grating detection (Chen, Li et al., 2016) or acuity-letter identification (Chung, Li, & Levi, 2012), or on binocular tasks that use dichoptic stimuli and involve e.g. discrimination of motion-direction (Black et al., 2011; R. F. Hess, Ding et al., 2016), phase-shift (Ding, Klein, & Levi, 2013; Kwon et al., 2014) or forms-identity (Eileen E. Birch et al., 2016; Kwon et al., 2015). Among our SED-tests that measured binocular combination, we found a good reliability and validity for tests #1 and #3; being this ‘paradigm’ a good index to measure vision improvement, these tests are potential instruments to assess the relative change, pre- vs post-, treatment for amblyopia.

On the other hand, consider *binocular rivalry* (Ooi & He, 2001; paradigm 2): the different contribution of the signal from each eye can be reversed, as perceptual dominance (i.e. the extent to which the FE dominates in rivalrous circumstances) has been shown to significantly change over the course of occlusion therapy (Lunghi, Morrone et al., 2016).

Potentially, further investigation might reveal the predictive role of results in binocular rivalry on the outcome of amblyopia treatment. However, our test #7 provided measures that were not particularly reliable, and it therefore might be unwise to base treatment assessment solely on rivalry-results, either obtained with our or similar tests.

Tests that involve *Flash-suppression*, e.g. form-discrimination emerging from the dynamic-noise pattern (delivering the target stimulus to one eye and a suppressive background to the other eye; Yang et al., 2010), could be used to investigate the unconscious side of suppression under rivalrous stimulation (Faivre, Berthet, & Kouider, 2014). To our knowledge, this process (potentially, paradigm 3) has not yet been specifically investigated in relation to amblyopia.

So far, we have considered adult-plasticity and key 'paradigms' that could reveal the effectiveness of a treatment in releasing-from-suppression amblyopic vision, or in other words paradigms that are key to quantify a re-balancing in SED.

We next consider, *what type of input is optimal for improving binocular vision in amblyopia*? The common principle in many binocular therapies for amblyopia is to progressively re-balance the visibility of dichoptic stimuli across the eyes (Ding & Levi, 2014; R. F. Hess & Thompson, 2015). The aim then is to balance SED, based on the notion that a degree of binocularity will be present but masked by the perceptual dominance of one-eye during perception. In fact, supra-threshold stimulation allows binocularity to occur normally also in amblyopic viewers (Mansouri et al., 2008). The most common method to train vision through re-balancing image visibility is to modulate the contrast of dichoptic stimuli (Ding & Sperling, 2006; C.-B. Huang et al., 2010). We have presented in section 2.8 the currently available methodologies. However (and this is especially the case) at high spatial frequencies, contrast sensitivity is significantly affected by reducing *luminance* (Campbell & Robson, 1968). In amblyopic viewers, placing a neutral density filter in front of one eye allows for modification of the unilateral mean luminance, and therefore contrast sensitivity (D. H. Baker et al., 2008), contrast-SED (Ding & Levi, 2014) and ultimately, binocular combination (J. Zhou, Jia, et al., 2013). One recent study on adults with normal vision highlights the role of *absolute* luminance of the inputs to both eyes in modifying SED (Yao et al., 2017), which resulted more incisive on binocular-phase-combination than the inter-ocular *relative* luminance (i.e. the ratio between monocular, absolute luminance values). Overall, changes in SED might critically depend on monocular changes in contrast sensitivity and/or luminance sensitivity. It has been proposed that severe and mild cases of amblyopia could be

assessed by focusing on sensitivity to *contrast*-modulated or *luminance*-modulated stimuli respectively since visual processes dealing with these features are differently sensitive to the depth of deficit (Chima et al., 2016). Considering results from the series of SED tests that we evaluated, we notice that test #3 was structured so that both contrast and luminance were modulated.

In summary, the practice of using a diversified set of stimuli for testing functional vision and the goal of achieving a complete and reliable vision assessment become interrelated concepts in moving the research on amblyopia forward, guiding the best practice in the treatment of this syndrome.

7.2.2 Assessing amblyopia at any age by measuring sensory eye dominance

The idea of an absolute critical-period for the recovery of visual function in amblyopia (circumscribed around the post-natal/early developmental phases of the visual system) is no longer a defensible idea. Multiple studies have demonstrated the effectiveness of amblyopia treatment also in adulthood (among others: T. Y. Gao et al., 2018; R. F. Hess, Babu, et al., 2014; J. Li, Spiegel et al., 2015; Liu & Zhang, 2017; Vedamurthy, Nahum, Huang, et al., 2015). Nevertheless, it is the case that early deprivation of visual stimulation causes more serious and longer-lasting loss of visual function, which ultimately results in more severe deficit (E. E. Birch, 2013). Thanks to key animal studies and retrospective studies of clinical records, we are understanding more about the ‘sensitive period’, along with the course of development of the visual system (for more details see respectively sections 2.4.1 and 1.4). It is now clear: i) when each visual ability is more vulnerable to abnormal stimulation, and ii) when we would expect specific behavioural effects to emerge. Broadly, we can distinguish a preliminary sensitive period for monocular vision, that goes from birth to ~4 months when acuity and contrast sensitivity are most vulnerable; and a period beginning a bit later, from ~18 months up to 7 years of age, that characterises the development of binocular vision and stereoacuity (Braddick & Atkinson, 2011). Treatment is potentially more effective during this relatively prolonged ‘recovery period’ and decreases with age (Daw, 1998), but should be offered even in adulthood (David K. Wallace et al., 2018). The pattern of amblyopia varies depending on the amblyogenic factor (McKee et al., 2003), so that according to the nature of the deficit (among other factors, such as age) we highlight the importance of a systematic evaluation of visual function in monitoring amblyopia treatment, especially considering the increasing interest to personalized therapies. In fact, being able ‘customize treatment’, could maximise the relative positive outcomes

while reducing treatment duration (and possibly costs), either if applied to patching-dose (Stewart et al., 2017) or binocular video-games/movies (not explored yet).

We suggest clinicians seriously consider the use of SED tests to provide a quantitative measure of the binocular deficit. The advantage of having a standardised set of tests to measure SED would be: i) characterising the pattern of amblyopia more in details, potentially defining specific pattern that could improve screening and direct a faster diagnosis, ii) define guideline to management and iii) allow a unique interpretation of research results in research, currently jeopardised by different investigative methods and interpretative criteria for the outcomes.

7.3 Future Directions

So far, randomized controlled trials of binocular alternative therapies have not demonstrated the superiority of these treatments compared to mainstream treatments for amblyopia. Anti-suppression gameplay (Tetris) did not provide a superior treatment of amblyopia than either 2-hrs of daily patching (J. M. Holmes et al., 2016) or placebo game-play (T. Y. Gao et al., 2018). Studies investigating monocular occlusion (Cotter, Edwards et al., 2007; Cotter et al., 2006; Stewart, Moseley, Fielder, et al., 2004) or binocular game-play (T.Y. Gao et al., 2018) demonstrated the importance of controlling for a period of full-wearing glasses (optical treatment, OT) prior to induction to any further treatment for amblyopia, to stabilize acuity and then evaluate the actual effects of the treatment (on top of those derived from optical correction). We only recruited children after minimum 16 weeks of OT, while earlier studies may have been contaminated by a lack of control for the OT. To further explore the impact of OT, new studies promise to shed light on the use of bifocal vs. single vision lenses during OT (pilot study plan: PEDIG-study number: ATS21) and on the impact of OT either alone- or in addition to-1hr/day of videogame (DigRush) play (PEDIG: multicentre RTC protocol-ATS20; trial ID: NCT02983552) – more details about both these studies are available at http://www.pedig.net/.

Further, we reasoned that one possible reason for the lack of consistency in results might be related to the way these treatments are evaluated, i.e. based on monocular acuity. Note that the above-mentioned studies implement a similar therapeutic principle that involves dichoptically presented games: progressively increase the initially-very-low contrast of the FE-image, until the game cannot be played anymore, as the signal from the AE-target-image, presented at a fixed high contrast is ‘suppressed’ by the noise from

the FE-image (i.e. until when the inter-ocular balance point is exceeded). This is an *individualised* procedure, where contrast levels used will vary greatly between participants. As well, the individual response might vary greatly, manifesting at different time-points throughout the course of the treatment. However, these therapies are usually evaluated based on pre-planned assessments, e.g. usually at 8 weeks distance from one another. During BBV, we did not change contrast throughout the treatment, which allowed us of not-having to consider the dynamics with which the individual response changed, on top of a change in the stimulation eliciting such a response. For consistency with other studies, we focused on acuity as the primary outcome and we based on that our early-termination criterion: at safe pre-planned time-point children were released from BBV, resulting in a different treatment duration between them (i.e. 8, 16, or 24 weeks long).

For decades, amblyopia *treatment-dose* and *-duration* have been fixed, making it easier to compare results and to realise, for example, that 2 vs. 6 hrs of daily patching are as effective as 6 vs. 12 hrs, respectively in case of moderate or severe cases of amblyopia (J. M. Holmes et al., 2003; Repka et al., 2003). From there, and thanks to the implementation of occlusion-dose-monitor devices, a model to calculate the optimal dose-rate for occlusion therapy has also been proposed (Loudon, Fronius et al., 2006), which incorporates important factors such as baseline acuity and age. For binocular treatments, the *exact time-course* of vision change is poorly documented and, when reported, results are unclear (T. Y. Gao et al., 2018). For example, it seems that the presence of strabismus might slow the rate of improvement, which varies between active or passive treatment modalities (Vedamurthy, Nahum, Huang, et al., 2015). Unclear is also the time-course of subsequent effect(s) on functional vision – although the possibility for the vision gain from binocular treatment to be maintained at one year from treatment cessation has been documented (S. L. Li, Jost et al., 2015). On this line, further investigation of novel binocular treatments for amblyopia should aim at implementing similar protocols between each other (e.g. similar definition of amblyopia and comparable scales to assess treatment results), so that the approaching new trials will help to solve the still unexplored aspect of dose-rate for these treatment methodologies.

7.4 Conclusion

Amblyopia is defined as a reduction of vision in one eye compared to the other, and the severity of deficit is typically related to the degree of imbalance (in acuity) between the two eyes. Results presented here support the importance of stimulating binocular vision, as a means to improve both binocular and monocular visual function.

Our proposed therapy is effective and safe, independently of the age, severity and type of amblyopia. Compared to other treatments, BBV incorporates a remotely administered measure of sensory-eye-dominance and compliance. For now, we cannot directly compare its efficacy to currently available therapies. However, implemented in a randomised controlled study, our treatment could offer a picture of the dose-response relationship following a binocular treatment, which is missing in the current literature.

Assessment of new therapies focuses on, monocular acuity, and, when measurable, stereo- acuity. The first ability, commonly tested using crowded charts, does not directly quantify the binocular deficit, which is *assumed* from the amount of inter-ocular difference in monocular results. The second, stereoacuity, is detectable only up to a certain level of sub-normal sensitivity, but usually results unmeasurable in severe amblyopes. We have highlighted the reliability and validity of testing inter-ocular imbalance using sensory-eye-dominance, assessed using a series of tests. Among these tests, a set of three appears particularly useful all of which assess recognition of dichoptic-letters (or symbols). Assessment of binocular vision using such a method would be an effective, quick and potentially convenient addition to tests used in the management of amblyopia. For example, the test could make clinicians aware of a sudden change in SED suggesting an increased risk of diplopia or ultimately of ‘reverse amblyopia’ (should the change persist). In term of practical aspects, the system we used for testing is not yet portable, but a number of suitable platforms (e.g. 3D video games) exist to support testing.

Overall, this thesis addressed our main questions as follows (note: this is drawn on the information summarised in this chapter, as well as on the Discussion and the other relevant sections in all the previous experimental chapters):

- *Are binocular treatments for amblyopia a valid alternative to traditional occlusion?*

Yes, they are a valid alternative that are demonstrably not inferior to traditional occlusion. Yet, there is no generalised clinical evidence for a superior therapeutic effect of one or another methodology. Binocular treatments tend to maximise compliance and be 'more enjoyable' for patients than traditional ones. They allow for a thorough stimulation of monocular *and* binocular visual abilities, with the potential of expanding the efficacy of a therapy to both aspects of vision, both typical of the amblyopia syndrome.

- *If so, how do they work?*

We speculate that monocular and binocular visual abilities improve following training due to a re-organisation of V1 receptive-fields' size and the relative neural network (*'re-mapping'*), yielding to the processing of binocularly balanced stimuli in higher visual areas, resulting in a reduced eye-dominance at the perceptual level.

- *Can we develop better tests for evaluating binocular visual function?*

We have presented and discussed a series of tests that have the advantage of being quick and easy to administer while offering reliable and valid results. Measuring the individual level of SED could support development of personalised-treatment regimens for amblyopia.

References

Achiron, L. R., Witkin, N., Primo, S., & Broocker, G. (1997). Contemporary management of aniseikonia. *Survey of Ophthalmology, 41*(4), 321-330. doi: http://dx.doi.org/10.1016/S0039-6257(96)00005-7

Adams, W. E., Leske, D. A., Hatt, S. R., & Holmes, J. M. (2009). Defining real change in measures of stereoacuity. *Ophthalmology, 116*(2), 281-285. doi: 10.1016/j.ophtha.2008.09.012

Adler, P., Scally, A. J., & Barrett, B. T. (2012). Test–retest variability of Randot stereoacuity measures gathered in an unselected sample of UK primary school children. *British Journal of Ophthalmology, 96*(5), 656-661. doi: 10.1136/bjophthalmol-2011-300729

Agervi, P., Kugelberg, U., Kugelberg, M., Simonsson, G., Fornander, M., & Zetterstrom, C. (2009). Treatment of anisometropic amblyopia with spectacles or in combination with translucent Bangerter filters. *Ophthalmology, 116*(8), 1475-1480.

Agrawal, R., Conner, I. P., Odom, J. V., Schwartz, T. L., & Mendola, J. D. (2006). Relating binocular and monocular vision in strabismic and anisometropic amblyopia. *Arch Ophthalmol, 124*(6), 844-850. doi: 10.1001/archopht.124.6.844

Alio, J. L., Wolter, N. V., Pinero, D. P., Amparo, F., Sari, E. S., Cankaya, C., & Laria, C. (2011). Pediatric refractive surgery and its role in the treatment of amblyopia: meta-analysis of the peer-reviewed literature. *J Refract Surg, 27*(5), 364-374. doi: 10.3928/1081597x-20100831-01

Allen, B., Spiegel, D. P., Thompson, B., Pestilli, F., & Rokers, B. (2015). Altered white matter in early visual pathways of humans with amblyopia. *Vision Res, 114*, 48-55. doi: http://dx.doi.org/10.1016/j.visres.2014.12.021

American Academy of Ophthalmology, & Pediatric Opthalmology/Strabismus Panel. (2012). *Preferred Practice Pattern. Guidelines. Amblyopia.*, San Francisco, CA. www.aao.org/ppp

Anketell, P. M., Saunders, K. J., & Little, J. A. (2013). Stereoacuity norms for school-age children using the Frisby stereotest. *J AAPOS, 17*(6), 582-587. doi: 10.1016/j.jaapos.2013.08.012

Anstice, N. S., & Thompson, B. (2014). The measurement of visual acuity in children: an evidence-based update. *Clinical and Experimental Optometry, 97*(1), 3-11. doi: 10.1111/cxo.12086

Anzai, A., Ohzawa, I., & Freeman, R. D. (1999). Neural mechanisms for processing binocular information II. Complex cells. *J Neurophysiol, 82*(2), 909-924.

Astle, A. T., Webb, B. S., & McGraw, P. V. (2011). The pattern of learned visual improvements in adult amblyopia. *Invest Ophthalmol Vis Sci, 52*(10), 7195-7204. doi: 10.1167/iovs.11-7584

Astle, W. F., Fawcett, S. L., Huang, P. T., Alewenah, O., & Ingram, A. (2008). Long-term outcomes of photorefractive keratectomy and laser-assisted subepithelial keratectomy in children. *J Cataract Refract Surg, 34*(3), 411-416. doi: 10.1016/j.jcrs.2007.10.027

Attebo, K., Mitchell, P., Cumming, R., Smith, W., Jolly, N., & Sparkes, R. (1998). Prevalence and causes of amblyopia in an adult population. *Ophthalmology, 105*(1), 154-159.

Awan, M., Proudlock, F. A., Grosvenor, D., Choudhuri, I., Sarvanananthan, N., & Gottlob, I. (2010). An audit of the outcome of amblyopia treatment: a retrospective analysis of 322 children. *Br J Ophthalmol, 94*(8), 1007-1011. doi: 10.1136/bjo.2008.154674

Babu, R. J., Clavagnier, S., Bobier, W. R., Thompson, B., & Hess, R. F. (2017). Regional Extent of Peripheral Suppression in Amblyopia. *Invest Ophthalmol Vis Sci, 58*(4), 2329-2340. doi: 10.1167/iovs.16-20012

Babu, R. J., Clavagnier, S. R., Bobier, W., Thompson, B., & Hess, R. F. (2013). The regional extent of suppression: strabismics versus nonstrabismics. *Invest Ophthalmol Vis Sci, 54*(10), 6585-6593. doi: 10.1167/iovs.12-11314

Baker, D. H., Kaestner, M., & Gouws, A. D. (2016). Measurement of crosstalk in stereoscopic display systems used for vision research. *Journal of Vision, 16*(15), 14. doi: 10.1167/16.15.14

Baker, D. H., Meese, T. S., & Hess, R. F. (2008). Contrast masking in strabismic amblyopia: attenuation, noise, interocular suppression and binocular summation. *Vision Res, 48*(15), 1625-1640. doi: 10.1016/j.visres.2008.04.017

Baker, D. H., Meese, T. S., Mansouri, B., & Hess, R. F. (2007). Binocular summation of contrast remains intact in strabismic amblyopia. *Invest Ophthalmol Vis Sci, 48*(11), 5332-5338. doi: 10.1167/iovs.07-0194

Baker, D. H., Meese, T. S., & Summers, R. J. (2007). Psychophysical evidence for two routes to suppression before binocular summation of signals in human vision. *Neuroscience, 146*(1), 435-448. doi: 10.1016/j.neuroscience.2007.01.030

Banks, M. S., & Salapatek, P. (1978). Acuity and contrast sensitivity in 1-, 2-, and 3-month-old human infants. *Invest Ophthalmol Vis Sci, 17*(4), 361-365.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *J Physiol, 193*(2), 327-342.

Barnes, G. R., Hess, R. F., Dumoulin, S. O., Achtman, R. L., & Pike, G. B. (2001). The cortical deficit in humans with strabismic amblyopia. *J Physiol, 533*(Pt 1), 281-297.

Barnes, G. R., Li, X., Thompson, B., Singh, K. D., Dumoulin, S. O., & Hess, R. F. (2010). Decreased gray matter concentration in the lateral geniculate nuclei in human amblyopes. *Invest Ophthalmol Vis Sci, 51*(3), 1432-1438. doi: 10.1167/iovs.09-3931

Barnet, A. B., Friedman, S. L., Weiss, I. P., Ohlrich, E. S., Shanks, B., & Lodge, A. (1980). VEP development in infancy and early childhood. A longitudinal study. *Electroencephalography and Clinical Neurophysiology, 49*(5), 476-489.

Baroncelli, L., Maffei, L., & Sale, A. (2011). New perspectives in amblyopia therapy on adults: a critical role for the excitatory/inhibitory balance. *Frontiers in cellular neuroscience, 5*.

Barrett, B. T., Bradley, A., & Candy, T. R. (2013). The relationship between anisometropia and amblyopia. *Progress in Retinal and Eye Research, 36*, 120-158. doi: 10.1016/j.preteyeres.2013.05.001

Bavelier, D., Levi, D. M., Li, R. W., Dan, Y., & Hensch, T. K. (2010). Removing brakes on adult brain plasticity: from molecular to behavioral interventions. *J Neurosci, 30*(45), 14964-14971.

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2001). *Neuroscience: Exploring the Brain* (C. Casco & L. Petrosini, Trans.): Masson.

Bedell, H. D., & Flom, M. C. (1981). Monocular spatial distortion in strabismic amblyopia. *Invest Ophthalmol Vis Sci, 20*(2), 263-268.

Bhola, R., Keech, R. V., Kutschke, P., Pfeifer, W., & Scott, W. E. (2006). Recurrence of Amblyopia after Occlusion Therapy. *Ophthalmology, 113*(11), 2097-2100.

Bi, H., Zhang, B., Tao, X., Harwerth, R. S., Smith, E. L., 3rd, & Chino, Y. M. (2011). Neuronal responses in visual area V2 (V2) of macaque monkeys with strabismic amblyopia. *Cereb Cortex, 21*(9), 2033-2045. doi: 10.1093/cercor/bhq272

Birch, E., Williams, C., Drover, J., Fu, V., Cheng, C., Northstone, K., . . . Adams, R. (2008). Randot Preschool Stereoacuity Test: normative data and validity. *J AAPOS, 12*(1), 23-26. doi: 10.1016/j.jaapos.2007.06.003

Birch, E. E. (2013). Amblyopia and binocular vision. *Progress in Retinal and Eye Research, 33*(0), 67-84.

Birch, E. E., Gwiazda, J., & Held, R. (1982). Stereoacuity development for crossed and uncrossed disparities in human infants. *Vision Res, 22*(5), 507-513.

Birch, E. E., Li, S. L., Jost, R. M., Morale, S. E., De La Cruz, A., Stager, D., Jr., . . . Stager, D. R., Sr. (2015). Binocular iPad treatment for amblyopia in preschool children. *J AAPOS, 19*(1), 6-11. doi: 10.1016/j.jaapos.2014.09.009

Birch, E. E., Morale, S. E., Jost, R. M., De La Cruz, A., Kelly, K. R., Wang, Y.-Z., & Bex, P. J. (2016). Assessing Suppression in Amblyopic Children With a Dichoptic Eye Chart. *Invest Ophthalmol Vis Sci, 57*(13), 5649-5654. doi: 10.1167/iovs.16-19986

Black, J. M., Hess, R. F., Cooperstock, J. R., To, L., & Thompson, B. (2012). The measurement and treatment of suppression in amblyopia. *J Vis Exp*(70), e3927. doi: 10.3791/3927

Black, J. M., Thompson, B., Maehara, G., & Hess, R. F. (2011). A compact clinical instrument for quantifying suppression. *Optom Vis Sci, 88*(2), E334-343. doi: 10.1097/OPX.0b013e318205a162

Blake, R., & Logothetis, N. (2002). Visual competition. *Nat Rev Neurosci, 3*(1), 13-21. doi: 10.1038/nrn701

Blakemore, C., Garey, L. J., & Vital-Durand, F. (1978). The physiological effects of monocular deprivation and their reversal in the monkey's visual cortex. *J Physiol, 283*(1), 223-262.

Blakemore, C., & Vital-Durand, F. (1986). Effects of visual deprivation on the development of the monkey's lateral geniculate nucleus. *J Physiol, 380*, 493-511.

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

Borchert, M., Tarczy-Hornoch, K., Cotter, S. A., Liu, N., Azen, S. P., & Varma, R. (2010). Anisometropia in Hispanic and African American Infants and Young Children: The Multi-Ethnic Pediatric Eye Disease Study. *Ophthalmology, 117*(1), 148-153.e141.

Bossi, M., Tailor, V. K., Anderson, E. J., Bex, P. J., Greenwood, J. A., Dahlmann-Noor, A., & Dakin, S. C. (2017). Binocular Therapy for Childhood Amblyopia Improves Vision Without Breaking Interocular Suppression. *Invest Ophthalmol Vis Sci, 58*(7), 3031-3043. doi: 10.1167/iovs.16-20913

Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., Hogg, R. E., & Mollon, J. D. (2015). A population study of binocular function. *Vision Res, 110*(Pt A), 34-50. doi: 10.1016/j.visres.2015.02.017

Braddick, O. (1979). Binocular Single Vision and Perceptual Processing. *Proceedings of the Royal Society of London. Series B, Biological Sciences, 204*(1157), 503-512.

Braddick, O. (1996). Binocularity in infancy. *Eye (Lond), 10 (Pt 2)*, 182-188. doi: 10.1038/eye.1996.45

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Res, 51*(13), 1588-1609.

Bradley, A., & Freeman, R. D. (1981). Contrast sensitivity in anisometropic amblyopia. *Invest Ophthalmol Vis Sci, 21*(3), 467-476.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Burian, H. M. (1943). Clinical significance of aniseikonia. *Arch Ophthalmol, 29*(1), 116-133. doi: 10.1001/archopht.1943.00880130136010

Burt, P., & Julesz, B. (1980). Modifications of the classical notion of Panum's fusional area. *Perception, 9*(6), 671-682.

Campbell, F. W., Hess, R. F., Watson, P. G., & Banks, R. (1978). Preliminary results of a physiologically based treatment of amblyopia. *Br J Ophthalmol, 62*(11), 748-755.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *J Physiol, 197*(3), 551-566.

Campos, E. C., & Enoch, J. M. (1980). Amount of aniseikonia compatible with fine binocular vision: some old and new concepts. *J Pediatr Ophthalmol Strabismus, 17*(1), 44-47.

Carkeet, A. (2015). Exact parametric confidence intervals for Bland-Altman limits of agreement. *Optom Vis Sci, 92*(3), e71-80. doi: 10.1097/OPX.0000000000000513

Carlton, J., & Kaltenthaler, E. (2011). Amblyopia and quality of life: a systematic review. *Eye (Lond), 25*(4), 403-413. doi: 10.1038/eye.2011.4

Casagrande, V. A. (1994). A third parallel visual pathway to primate area V1. *Trends Neurosci, 17*(7), 305-310.

Cattaneo, Z., Vecchi, T., Monegato, M., Pece, A., Merabet, L. B., & Carbon, C. C. (2013). Strabismic amblyopia affects relational but not featural and Gestalt processing of faces. *Vision Res, 80*, 19-30. doi: 10.1016/j.visres.2013.01.007

Chadnova, E., Reynaud, A., Clavagnier, S., & Hess, R. F. (2017). Short-term monocular occlusion produces changes in ocular dominance by a reciprocal modulation of interocular inhibition. *Sci Rep, 7*.

Chapman, B., & Stryker, M. P. (1993). Development of orientation selectivity in ferret visual cortex and effects of deprivation. *The Journal of Neuroscience, 13*(12), 5251-5262.

Chen, Z., Li, J., Liu, J., Cai, X., Yuan, J., Deng, D., & Yu, M. (2016). Monocular perceptual learning of contrast detection facilitates binocular combination in adults with anisometropic amblyopia. *Sci Rep, 6*, 20187. doi: 10.1038/srep20187

Chima, A. S., Formankiewicz, M. A., & Waugh, S. J. (2016). Interocular suppression patterns in binocularly abnormal observers using luminance- and contrast-modulated noise stimuli. *Journal of Vision, 16*(10), 20-20. doi: 10.1167/16.10.20

Christakis, D. A., Ebel, B. E., Rivara, F. P., & Zimmerman, F. J. (2004). Television, video, and computer game usage in children under 11 years of age. *J Pediatr, 145*(5), 652-656.
Chua, B., & Mitchell, P. (2004). Consequences of amblyopia on education, occupation, and long term vision loss. *Br J Ophthalmol, 88*(9), 1119-1121. doi: 10.1136/bjo.2004.041863
Chung, S. T., Li, R. W., & Levi, D. M. (2012). Learning to identify near-acuity letters, either with or without flankers, results in improved letter size and spacing limits in adults with amblyopia. *PLoS One, 7*(4), e35829. doi: 10.1371/journal.pone.0035829
Ciuffreda, K. J., Levi, D. M., & Selenow, A. (1991). *Amblyopia: Basic and Clinical Aspects*: Butterworth-Heinemann.
Clavagnier, S., Dumoulin, S. O., & Hess, R. F. (2015). Is the Cortical Deficit in Amblyopia Due to Reduced Cortical Magnification, Loss of Neural Resolution, or Neural Disorganization? *J Neurosci, 35*(44), 14740-14755. doi: 10.1523/jneurosci.1101-15.2015
Clavagnier, S., Thompson, B., & Hess, R. F. (2013). Long Lasting Effects of Daily Theta Burst rTMS Sessions in the Human Amblyopic Cortex. *Brain Stimul*. doi: 10.1016/j.brs.2013.04.002
Cleary, M., Moody, A. D., Buchanan, A., Stewart, H., & Dutton, G. N. (2009). Assessment of a computer-based treatment for older amblyopes: the Glasgow Pilot Study. *Eye (Lond), 23*(1), 124-131. doi: 10.1038/sj.eye.6702977
Conner, I. P., Odom, J. V., Schwartz, T. L., & Mendola, J. D. (2007). Monocular activation of V1 and V2 in amblyopic adults measured with functional magnetic resonance imaging. *J AAPOS, 11*(4), 341-350. doi: 10.1016/j.jaapos.2007.01.119
Cormack, L. K., Scott, B. S., & Schor, C. M. (1993). Disparity-tuned channels of the human visual system. *Vis Neurosci, 10*, 585-596.
Cotter, S. A., Cyert, L. A., Miller, J. M., & Quinn, G. E. (2015). Vision screening for children 36 to <72 months: recommended practices. *Optom Vis Sci, 92*(1), 6-16. doi: 10.1097/opx.0000000000000429
Cotter, S. A., Edwards, A. R., Arnold, R. W., Astle, W. F., Barnhardt, C. N., Beck, R. W., . . . Weise, K. K. (2007). Treatment of strabismic amblyopia with refractive correction. *Am J Ophthalmol, 143*(6), 1060-1063. doi: 10.1016/j.ajo.2007.02.029
Cotter, S. A., Edwards, A. R., Wallace, D. K., Beck, R. W., Arnold, R. W., Astle, W. F., . . . Weise, K. K. (2006). Treatment of anisometropic amblyopia in children with refractive correction. *Ophthalmology, 113*(6), 895-903.
Cotter, S. A., Foster, N. C., Holmes, J. M., Melia, B. M., Wallace, D. K., Repka, M. X., . . . Suh, D. W. (2012). Optical treatment of strabismic and combined strabismic-anisometropic amblyopia. *Ophthalmology, 119*(1), 150-158. doi: 10.1016/j.ophtha.2011.06.043
Crawford, L. C. J., & Griffiths, H. J. (2015). The repeatability of the Sbisa bar for testing density of suppression. *British and Irish Orthoptic Journal, 12*, 35-40.
Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annu Rev Neurosci, 24*, 203-238. doi: 10.1146/annurev.neuro.24.1.203
Dacey, D., Packer, O. S., Diller, L., Brainard, D., Peterson, B., & Lee, B. (2000). Center surround receptive field structure of cone bipolar cells in primate retina. *Vision Res, 40*(14), 1801-1811.
Daw, N. W. (1998). Critical periods and amblyopia. *Arch Ophthalmol, 116*(4), 502-505. doi: 10.1001/archopht.116.4.502
De Valois, R. L., Yund, W. E., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res, 22*(5), 531-544.
Demer, J. L., Grafton, S., Marg, E., Mazziotta, J. C., & Nuwer, M. (1997). Positron-emission tomographic study of human amblyopia with use of defined visual stimuli. *J AAPOS, 1*(3), 158-171.
Demer, J. L., von Noorden, G. K., Volkow, N. D., & Gould, K. L. (1988). Imaging of cerebral blood flow and metabolism in amblyopia by positron emission tomography. *Am J Ophthalmol, 105*(4), 337-347.
Dieter, K. C., Sy, J. L., & Blake, R. (2016). Individual differences in sensory eye dominance reflected in the dynamics of binocular rivalry. *Vision Res*. doi: 10.1016/j.visres.2016.09.014
Ding, J., Klein, S. A., & Levi, D. M. (2013). Binocular Combination in Abnormal Binocular Vision. *Journal of Vision, 13(2)*(14), 1-31. doi: 10.1167/13.2.14
Ding, J., & Levi, D. M. (2011). Recovery of stereopsis through perceptual learning in human adults with abnormal binocular vision. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1105183108

Ding, J., & Levi, D. M. (2014). Rebalancing binocular vision in amblyopia. *Ophthalmic and Physiological Optics*, n/a-n/a. doi: 10.1111/opo.12115

Ding, J., & Sperling, G. (2006). A gain-control theory of binocular combination. *Proc Natl Acad Sci U S A, 103*(4), 1141-1146. doi: 10.1073/pnas.0509629103

Dobson, V. (1980). Behavioral tests of visual acuity in infants. *Int Ophthalmol Clin, 20*(1), 233-250.

Dorr, M., Elze, T., Wang, H., Lu, Z. L., Bex, P. J., & Lesmes, L. (2017). *New precision metrics for contrast sensitivity testing*.

Dorr, M., Lesmes, L. A., Lu, Z.-L., & Bex, P. J. (2013). Rapid and Reliable Assessment of the Contrast Sensitivity Function on an iPad. *Invest Ophthalmol Vis Sci, 54*(12), 7266-7273. doi: 10.1167/iovs.13-11743

Duan, Y., Norcia, A. M., Yeatman, J. D., & Mezer, A. (2015). The Structural Properties of Major White Matter Tracts in Strabismic Amblyopia. *Invest Ophthalmol Vis Sci, 56*(9), 5152-5160. doi: 10.1167/iovs.15-17097

Duffy, K. R., & Mitchell, D. E. (2013). Darkness alter maturation of visual cortex and promotes fast recovery from monocular deprivation. *Curr Biol, 23*, 382-386. doi: 10.1016/j.cub.2013.01.017

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage, 39*(2), 647-660. doi: 10.1016/j.neuroimage.2007.09.034

Eastgate, R. M., Griffiths, G. D., Waddingham, P. E., Moody, A. D., Butler, T. K. H., Cobb, S. V., . . . Brown, S. M. (2005). Modified virtual reality technology for treatment of amblyopia. *Eye, 20*(3), 370-374. doi: 10.1038/sj.eye.6701882

Ehrenstein, W. H., Arnold-Schulz-Gahmen, B. E., & Jaschinski, W. (2005). Eye preference within the context of binocular functions. *Graefe's Archive for Clinical and Experimental Ophthalmology, 243*(9), 926-932.

Erchova, I., Vasalauskaite, A., Longo, V., & Sengpiel, F. (2017). Enhancement of visual cortex plasticity by dark exposure. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1715), 20160159. doi: 10.1098/rstb.2016.0159

Erdem, E., Cinar, G. Y., Somer, D., Demir, N., Burcu, A., & Ornek, F. (2011). Eye patching as a treatment for amblyopia in children aged 10-16 years. *Jpn J Ophthalmol, 55*(4), 389-395. doi: 10.1007/s10384-011-0029-z

Espinosa, J. S., & Stryker, M. P. (2012). Development and plasticity of the primary visual cortex. *Neuron, 75*(2), 230-249. doi: 10.1016/j.neuron.2012.06.009

Evans, B. J. W., Yu, C. S., Massa, E., & Mathews, J. E. (2011). Randomised controlled trial of intermittent photopic stimulation for treating amblyopia in older children and adults. *Ophthalmic Physiol Opt, 31*, 56-68. doi: 10.1111/j.1475-1313.2010.00801.x

Faivre, N., Berthet, V., & Kouider, S. (2014). Sustained invisibility through crowding and continuous flash suppression: a comparative review. *Frontiers in Psychology, 5*(475). doi: 10.3389/fpsyg.2014.00475

Fantz, R. L. (1958). Pattern vision in young infants. *The Psychological Record*.

Farivar, R., Zhou, J., Huang, Y., Feng, L., Zhou, Y., & Hess, R. F. (2017). Two cortical deficits underlie amblyopia: A multi-focal fMRI analysis. *Neuroimage*.

Felius, J., Chandler, D. L., Holmes, J. M., Chu, R. H., Cole, S. R., Hill, M., . . . Wallace, D. K. (2010). Evaluating the Burden of Amblyopia Treatment from the Parent and Child's Perspective. *J AAPOS, 14*(5), 389-395. doi: 10.1016/j.jaapos.2010.07.009

Ferris, F. L., 3rd, Kassoff, A., Bresnick, G. H., & Bailey, I. (1982). New visual acuity charts for clinical research. *Am J Ophthalmol, 94*(1), 91-96.

Freeman, R. D., & Ohzawa, I. (1990). On the neurophysiological organization of binocular vision. *Vision Res, 30*(11), 1661-1676.

Fricke, T. R., & Siderov, J. (1997). Stereopsis, stereotests, and their relation to vision screening and clinical practice. *Clinical and Experimental Optometry, 80*(5), 165-172. doi: 10.1111/j.1444-0938.1997.tb04876.x

Friedman, D. S., Repka, M. X., Katz, J., Giordano, L., Ibironke, J., Hawse, P., & Tielsch, J. M. (2009). Prevalence of amblyopia and strabismus in white and African American children aged 6 through 71 months the Baltimore Pediatric Eye Disease Study. *Ophthalmology, 116*(11), 2128-2134.e2121-2122. doi: 10.1016/j.ophtha.2009.04.034

Frisby, J. P., Davis, H., & McMorrow, K. (1996). An improved training procedure as a precursor to testing young children with the Frisby Stereotest. *Eye (Lond), 10 (Pt 2)*, 286-290. doi: 10.1038/eye.1996.60

Fronius, M., Sireteanu, R., & Zubcov, A. (2004). Deficits of spatial localization in children with strabismic amblyopia. *Graefes Arch Clin Exp Ophthalmol, 242*(10), 827-839. doi: 10.1007/s00417-004-0936-5

Gao, T. Y., Anstice, A., Babu, R. J., Black, J., Bobier, W., Dai, S., . . . Thompson, B. (2018). Optical treatment of amblyopia in older children and adults is essential prior to enrolment in a clinical trial. *Ophthalmic and Physiological Optics*. doi: 10.1111/opo.12437

Gao, T. Y., Guo, C. X., Babu, R. J., & et al. (2018). Effectiveness of a binocular video game vs placebo video game for improving visual functions in older children, teenagers, and adults with amblyopia: A randomized clinical trial. *JAMA Ophthalmol*. doi: 10.1001/jamaophthalmol.2017.6090

Gilbert, C. D., & Wiesel, T. N. (1992). Receptive field dynamics in adult primary visual cortex. *Nature, 356*(6365), 150-152.

Giordano, L., Friedman, D. S., Repka, M. X., Katz, J., Ibironke, J., Hawes, P., & Tielsch, J. M. (2009). Prevalence of refractive error among preschool children in an urban population: the Baltimore Pediatric Eye Disease Study. *Ophthalmology, 116*(4), 739-746, 746.e731-734. doi: 10.1016/j.ophtha.2008.12.030

Glass, L. (1969). Moire effect from random dots. *Nature, 223*(5206), 578-580.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci, 15*(1), 20-25.

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. *Invest Ophthalmol Vis Sci, 53*(12), 7655-7665. doi: 10.1167/iovs.12-10313

Hainline, B. C., Sprunger, D. C., Plager, D. A., Neely, D. E., & Guess, M. G. (2009). Reverse amblyopia with atropine treatment. *Binocul Vis Strabismus Q, 24*(1), 25-31.

Hamm, L., Chen, Z., Li, J., Black, J., Dai, S., Yuan, J., . . . Thompson, B. (2017). Interocular suppression in children with deprivation amblyopia. *Vision Res, 133*, 112-120. doi: 10.1016/j.visres.2017.01.004

Harrad, R., Sengpiel, F., & Blakemore, C. (1996). Physiology of suppression in strabismic amblyopia. *Br J Ophthalmol, 80*(4), 373-377.

Hartline, H. K. (1938). *The response of single optic nerve fibers of the vertebrate eye to illumination of the retina* (Vol. 121).

Harvey, B. M., & Dumoulin, S. O. (2011). The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. *The Journal of Neuroscience, 31*(38), 13604-13612.

Harwerth, R. S., Smith, E. L., 3rd, Duncan, G. C., Crawford, M. L., & Von Noorden, G. K. (1986). Multiple sensitive periods in the development of the primate visual system. *Science, 232*(4747), 235-238.

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behav Res Methods*. doi: 10.3758/s13428-017-0935-1

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nat Rev Neurosci, 6*(11), 877-888. doi: 10.1038/nrn1787

Herbison, N., Cobb, S., Gregson, R., Ash, I., Eastgate, R., Purdy, J., . . . Foss, A. (2013). Interactive binocular treatment (I-BiT) for amblyopia: results of a pilot study of 3D shutter glasses system. *Eye (Lond), 27*(9), 1077-1083. doi: 10.1038/eye.2013.113

Herbison, N., MacKeith, D., Vivian, A., Purdy, J., Fakis, A., Ash, I. M., . . . Foss, A. J. (2016). Randomised controlled trial of video clips and interactive games to improve vision in children with amblyopia using the I-BiT system. *British Journal of Ophthalmology*. doi: 10.1136/bjophthalmol-2015-307798

Heron, S., & Lages, M. (2012). Screening and sampling in studies of binocular vision. *Vision Res, 62*, 228-234. doi: 10.1016/j.visres.2012.04.012

Hess, R. F., & Anderson, S. J. (1993). Motion sensitivity and spatial undersampling in amblyopia. *Vision Res, 33*(7), 881-896.

Hess, R. F., Babu, R. J., Clavagnier, S., Black, J., Bobier, W., & Thompson, B. (2014). The iPod binocular home-based treatment for amblyopia in adults: efficacy and compliance. *Clin Exp Optom, 97*(5), 389-398. doi: 10.1111/cxo.12192

Hess, R. F., & Demanins, R. (1998). Contour integration in anisometropic amblyopia. *Vision Res, 38*(6), 889-894.

Hess, R. F., Ding, R., Clavagnier, S., Liu, C., Guo, C., Viner, C., . . . Zhou, J. (2016). A Robust and Reliable Test to Measure Stereopsis in the Clinic. *Invest Ophthalmol Vis Sci, 57*(3), 798-804. doi: 10.1167/iovs.15-18690

Hess, R. F., & Field, D. J. (1994). Is the spatial deficit in strabismic amblyopia due to loss of cells or an uncalibrated disarray of cells? *Vision Res, 34*(24), 3397-3406.

Hess, R. F., & Howell, E. R. (1977). The threshold contrast sensitivity function in strabismic amblyopia: evidence for a two type classification. *Vision Res, 17*(9), 1049-1055.

Hess, R. F., Mansouri, B., & Thompson, B. (2010a). A binocular approach to treating amblyopia: antisuppression therapy. *Optom Vis Sci, 87*(9), 697-704. doi: 10.1097/OPX.0b013e3181ea18e9

Hess, R. F., Mansouri, B., & Thompson, B. (2010b). A new binocular approach to the treatment of amblyopia in adults well beyond the critical period of visual development. *Restor Neurol Neurosci, 28*(6), 793-802. doi: 10.3233/rnn-2010-0550

Hess, R. F., Mansouri, B., & Thompson, B. (2011). Restoration of binocular vision in amblyopia. *Strabismus, 19*(3), 110-118. doi: 10.3109/09273972.2011.600418

Hess, R. F., McIlhagga, W., & Field, D. J. (1997). Contour integration in strabismic amblyopia: the sufficiency of an explanation based on positional uncertainty. *Vision Res, 37*(22), 3145-3161.

Hess, R. F., & Thompson, B. (2013). New insights into amblyopia: binocular therapy and noninvasive brain stimulation. *J AAPOS, 17*(1), 89-93. doi: 10.1016/j.jaapos.2012.10.018

Hess, R. F., & Thompson, B. (2015). Amblyopia and the binocular approach to its therapy. *Vision Res*. doi: 10.1016/j.visres.2015.02.009

Hess, R. F., Thompson, B., & Baker, D. H. (2014). Binocular vision in amblyopia: structure, suppression and plasticity. *Ophthalmic Physiol Opt, 34*(2), 146-162. doi: 10.1111/opo.12123

Hess, R. F., Thompson, B., Black, J. M., Machara, G., Zhang, P., Bobier, W. R., & Cooperstock, J. (2012). An iPod treatment of amblyopia: an updated binocular approach. *Optometry, 83*(2), 87-94.

Hess, R. F., To, L., Zhou, J., Wang, G., & Cooperstock, J. R. (2015). Stereo Vision: The Haves and Have-Nots. *i-Perception, 6*(3). doi: 10.1177/2041669515593028

Ho, C. S., & Giaschi, D. E. (2009). Low- and high-level motion perception deficits in anisometropic and strabismic amblyopia: evidence from fMRI. *Vision Res, 49*(24), 2891-2901. doi: 10.1016/j.visres.2009.07.012

Holmes, J., Beck, R., Kraker, R., Astle, W., Birch, E., Cole, S., . . . Scheiman, M. (2004). Risk of amblyopia recurrence after cessation of treatment. *J AAPOS, 8*(5), 420-428. doi: 10.1016/s1091853104001612

Holmes, J. M., Beck, R. W., Repka, M. X., Leske, D. A., Kraker, R. T., Blair, R. C., . . . Miller, J. M. (2001). The amblyopia treatment study visual acuity testing protocol. *Arch Ophthalmol, 119*(9), 1345-1353.

Holmes, J. M., & Clarke, M. P. (2006). Amblyopia. *Lancet, 367*(9519), 1343-1351. doi: 10.1016/S0140-6736(06)68581-4

Holmes, J. M., Kraker, R. T., Beck, R. W., Birch, E. E., Cotter, S. A., Everett, D. F., . . . Wallace, D. K. (2003). A randomized trial of prescribed patching regimens for treatment of severe amblyopia in children. *Ophthalmology, 110*(11), 2075-2087.

Holmes, J. M., Lazar, E. L., Melia, B. M., Astle, W. F., Dagi, L. R., Donahue, S. P., . . . Weise, K. K. (2011). Effect of age on response to amblyopia treatment in children. *Arch Ophthalmol, 129*(11), 1451-1457. doi: 10.1001/archophthalmol.2011.179

Holmes, J. M., Manh, V. M., Lazar, E. L., & et al. (2016). Effect of a binocular ipad game vs part-time patching in children aged 5 to 12 years with amblyopia: A randomized clinical trial. *JAMA Ophthalmol*. doi: 10.1001/jamaophthalmol.2016.4262

Holopigian, K., Blake, R., & Greenwald, M. J. (1986). Selective losses in binocular vision in anisometropic amblyopes. *Vision Res, 26*(4), 621-630.

Hou, F., Huang, C.-B., Lesmes, L., Feng, L.-X., Tao, L., Zhou, Y.-F., & Lu, Z.-L. (2010). qCSF in Clinical Application: Efficient Characterization and Classification of Contrast Sensitivity Functions in Amblyopia. *Invest Ophthalmol Vis Sci, 51*(10), 5365-5377. doi: 10.1167/iovs.10-5468

Hrisos, S., Clarke, M. P., & Wright, C. M. (2004). The emotional impact of amblyopia treatment in preschool children: randomized controlled trial. *Ophthalmology, 111*(8), 1550-1556. doi: 10.1016/j.ophtha.2003.12.059

Huang, C.-B., Zhou, J., Zhou, Y., & Lu, Z.-L. (2010). Contrast and phase combination in binocular vision. *PLoS One, 5*(12), e15075.
Huang, C. B., Zhou, J., Lu, Z. L., Feng, L., & Zhou, Y. (2009). Binocular combination in anisometropic amblyopia. *J Vis, 9*(3), 17 11-16. doi: 10.1167/9.3.17
Huang, C. B., Zhou, J., Lu, Z. L., & Zhou, Y. (2011). Deficient Binocular Combination Reveals Mechanisms of Anisometropic Amblyopia: Signal Attenuation and Interocular Inhibition. *Journal of Vision, 11*(6), 10.1167/1111.1166.1164 1164. doi: 10.1167/11.6.4
Huang, P. C., Baker, D. H., & Hess, R. F. (2012). Interocular suppression in normal and amblyopic vision: spatio-temporal properties. *J Vis, 12*(11). doi: 10.1167/12.11.29
Hubel, D. H. (1982). Exploration of the primary visual cortex, 1955-78. *Nature, 299*(5883), 515-524.
Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol, 160*(1), 106-154.102.
Hubel, D. H., & Wiesel, T. N. (1963). Receptive fields of cells in Striate Cortex of very young, visually inexperienced kittens. *J Neurophysiol, 26*, 994-1002.
Hubel, D. H., & Wiesel, T. N. (1965a). Binocular interaction in striate cortex of kittens reared with artificial squint. *J Neurophysiol, 28*(6), 1041-1059.
Hubel, D. H., & Wiesel, T. N. (1965b). Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *J Neurophysiol, 28*, 229-289.
Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol, 195*(1), 215-243.
Hubel, D. H., & Wiesel, T. N. (1970a). Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature, 225*, 41-42.
Hubel, D. H., & Wiesel, T. N. (1970b). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *J Physiol, 206*(2), 419-436.
Hubel, D. H., Wiesel, T. N., & LeVay, S. (1976). Functional architecture of area 17 in normal and monocularly deprived macaque monkeys. *Cold Spring Harb Symp Quant Biol, 40*, 581-589.
Hubel, D. H., Wiesel, T. N., & LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Philos Trans R Soc Lond B Biol Sci, 278*(961), 377-409.
Husk, J. S., Farivar, R., & Hess, R. F. (2012). Amblyopic deficits in processing structure-from-motion. *J Vis, 12*(4). doi: 10.1167/12.4.4
Husk, J. S., & Hess, R. F. (2013). Global processing of orientation in amblyopia. *Vision Res, 82*, 22-30. doi: 10.1016/j.visres.2013.02.005
Hussain, Z., Astle, A. T., Webb, B. S., & McGraw, P. V. (2014). The challenges of developing a contrast-based video game for treatment of amblyopia. *Front Psychol, 5*. doi: 10.3389/fpsyg.2014.01210
Hussain, Z., Svensson, C.-M., Besle, J., Webb, B. S., Barrett, B. T., & McGraw, P. V. (2015). Estimation of cortical magnification from positional error in normally sighted and amblyopic subjects. *Journal of Vision, 15*(2), 25. doi: 10.1167/15.2.25
Hussain, Z., Webb, B. S., Astle, A. T., & McGraw, P. V. (2012). Perceptual learning reduces crowding in amblyopia and in the normal periphery. *J Neurosci, 32*(2), 474-480. doi: 10.1523/jneurosci.3845-11.2012
Ikeda, H. (1980). Visual acuity, its development and amblyopia. *J R Soc Med, 73*(8), 546.
Jampolsky, A. (1955). Characteristics of suppression in strabismus. *AMA Arch Ophthalmol, 54*(5), 683-696.
Johansson, J., Seimyr, G. Ö., & Pansell, T. (2015). Eye dominance in binocular viewing conditions. *Journal of Vision, 15*(9), 21-21.
Joly, O., & Franko, E. (2014). Neuroimaging of amblyopia and binocular vision: a review. *Front Integr Neurosci, 8*, 62. doi: 10.3389/fnint.2014.00062
Julesz, B. (1971). *Foundations of Cyclopean Perception*: University of Chicago Press.
Kehrein, S., Kohnen, T., & Fronius, M. (2016). Dynamics of Interocular Suppression in Amblyopic Children during Electronically Monitored Occlusion Therapy: First Insight. *Strabismus, 24*(2), 51-62. doi: 10.3109/09273972.2016.1170047
Kelly, K. R., Jost, R. M., Dao, L., Beauchamp, C. L., Leffler, J. N., & Birch, E. E. (2016). Binocular iPad Game vs Patching for Treatment of Amblyopia in Children: A Randomized Clinical Trial. *JAMA Ophthalmol, 134*(12), 1402-1408. doi: 10.1001/jamaophthalmol.2016.4224
Kelly, K. R., Jost, R. M., De La Cruz, A., & Birch, E. E. (2015). Amblyopic children read more slowly than controls under natural, binocular reading conditions. *J AAPOS, 19*(6), 515-520. doi: 10.1016/j.jaapos.2015.09.002

Kelly, K. R., Jost, R. M., Wang, Y.-Z., Dao, L., Beauchamp, C. L., Leffler, J. N., & Birch, E. E. (2018). Improved Binocular Outcomes Following Binocular Treatment for Childhood Amblyopia. *Invest Ophthalmol Vis Sci, 59*(3), 1221-1228. doi: 10.1167/iovs.17-23235

Khan, A. Z., & Crawford, J. D. (2001). Ocular dominance reverses as a function of horizontal gaze angle. *Vision Research, 41*(14), 1743-1748. doi: https://doi.org/10.1016/S0042-6989(01)00079-7

Kiorpes, L., Kiper, D. C., O'Keefe, L. P., Cavanaugh, J. R., & Movshon, J. A. (1998). Neuronal correlates of amblyopia in the visual cortex of macaque monkeys with experimental strabismus and anisometropia. *J Neurosci, 18*(16), 6411-6424.

Kishimoto, F., Fujii, C., Shira, Y., Hasebe, K., Hamasaki, I., & Ohtsuki, H. (2014). Outcome of conventional treatment for adult amblyopia. *Jpn J Ophthalmol, 58*(1), 26-32. doi: 10.1007/s10384-013-0279-z

Knox, P. J., Simmers, A. J., Gray, L. S., & Cleary, M. (2012). An exploratory study: prolonged periods of binocular stimulation can provide an effective treatment for childhood amblyopia. *Invest Ophthalmol Vis Sci, 53*(2), 817-824. doi: 10.1167/iovs.11-8219

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155-163. doi: 10.1016/j.jcm.2016.02.012

Kuffler, S. W. (1953). *Discharge patterns and functional organisation of mammalian retina* (Vol. 16).

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *J Physiol, 232*(1), 149-162. doi: 10.1113/jphysiol.1973.sp010261

Kwon, M., Lu, Z. L., Miller, A., Kazlas, M., Hunter, D. G., & Bex, P. J. (2014). Assessing Binocular Interaction in Amblyopia and Its Clinical Feasibility. *PLoS One, 9*(6), e100156. doi: 10.1371/journal.pone.0100156

Kwon, M., Wiecek, E., Dakin, S. C., & Bex, P. J. (2015). Spatial-frequency dependent binocular imbalance in amblyopia. *Sci Rep, 5*, 17181. doi: 10.1038/srep17181

Lai, X. J., Alexander, J., He, M., Yang, Z., & Suttle, C. (2011). Visual functions and interocular interactions in anisometropic children with and without amblyopia. *Invest Ophthalmol Vis Sci, 52*(9), 6849-6859. doi: 10.1167/iovs.10-6755

Lalor, S. J. H., Formankiewicz, M. A., & Waugh, S. J. (2016). Crowding and visual acuity measured in adults using paediatric test letters, pictures and symbols. *Vision Res, 121*, 31-38. doi: https://doi.org/10.1016/j.visres.2016.01.007

Legge, G. E. (1979). Spatial frequency masking in human vision: binocular interactions. *Journal of the Optical Society of America, 69*(6), 838-847. doi: 10.1364/JOSA.69.000838

LeVay, S., Hubel, D. H., & Wiesel, T. N. (1975). The pattern of ocular dominance columns in macaque visual cortex revealed by a reduced silver stain. *The Journal of Comparative Neurology, 159*(4), 559-575. doi: 10.1002/cne.901590408

Levi, D. M. (1994). Pathophysiology of binocular vision and amblyopia. *Curr Opin Ophthalmol, 5*(5), 3-10.

Levi, D. M. (2008). Crowding--an essential bottleneck for object recognition: a mini-review. *Vision Res, 48*(5), 635-654. doi: 10.1016/j.visres.2007.12.009

Levi, D. M., & Harwerth, R. S. (1977). Spatio-temporal interactions in anisometropic and strabismic amblyopia. *Invest Ophthalmol Vis Sci, 16*(1), 90-95.

Levi, D. M., & Klein, S. A. (1983). Spatial localization in normal and amblyopic vision. *Vision Res, 23*(10), 1005-1017.

Levi, D. M., & Klein, S. A. (1985). Vernier acuity, crowding and amblyopia. *Vision Res, 25*(7), 979-991.

Levi, D. M., Klein, S. A., & Chen, I. (2008). What limits performance in the amblyopic visual system: seeing signals in noise with an amblyopic brain. *J Vis, 8*(4), 1.1-23. doi: 10.1167/8.4.1

Levi, D. M., Klein, S. A., & Hariharan, S. (2002). Suppressive and facilitatory spatial interactions in foveal vision: foveal crowding is simple contrast masking. *J Vis, 2*(2), 140-166. doi: 10.1167/2.2.2

Levi, D. M., Klein, S. A., & Sharma, V. (1999). Position jitter and undersampling in pattern perception. *Vision Res, 39*(3), 445-465.

Levi, D. M., & Li, R. W. (2009). Perceptual learning as a potential treatment for amblyopia: a mini-review. *Vision Res, 49*(21), 2535-2549. doi: 10.1016/j.visres.2009.02.010

Levi, D. M., McKee, S. P., & Movshon, J. A. (2011). Visual deficits in anisometropia. *Vision Res, 51*(1), 48-57. doi: 10.1016/j.visres.2010.09.029

Lewis, T. L., & Maurer, D. (2005). Multiple sensitive periods in human visual development: evidence from visually deprived children. *Dev Psychobiol, 46*(3), 163-183.

Li, J., Hess, R. F., Chan, L. Y., Deng, D., Yang, X., Chen, X., . . . Thompson, B. (2013). Quantitative measurement of interocular suppression in anisometropic amblyopia: a case-control study. *Ophthalmology, 120*(8), 1672-1680. doi: 10.1016/j.ophtha.2013.01.048

Li, J., Lam, C. S., Yu, M., Hess, R. F., Chan, L. Y., Maehara, G., . . . Thompson, B. (2010). Quantifying sensory eye dominance in the normal visual system: a new technique and insights into variation across traditional tests. *Invest Ophthalmol Vis Sci, 51*(12), 6875-6881. doi: 10.1167/iovs.10-5549

Li, J., Spiegel, D. P., Hess, R. F., Chen, Z., Chan, L. Y., Deng, D., . . . Thompson, B. (2015). Dichoptic training improves contrast sensitivity in adults with amblyopia. *Vision Res*. doi: 10.1016/j.visres.2015.01.017

Li, J., Thompson, B., Deng, D., Chan, L. Y., Yu, M., & Hess, R. F. (2013). Dichoptic training enables the adult amblyopic brain to learn. *Curr Biol, 23*(8), R308-309. doi: 10.1016/j.cub.2013.01.059

Li, J., Thompson, B., Lam, C. S., Deng, D., Chan, L. Y., Maehara, G., . . . Hess, R. F. (2011). The role of suppression in amblyopia. *Invest Ophthalmol Vis Sci, 52*(7), 4169-4176. doi: 10.1167/iovs.11-7233

Li, S. L., Jost, R. M., Morale, S. E., & et al. (2015). Binocular ipad treatment of amblyopia for lasting improvement of visual acuity. *JAMA Ophthalmol, 133*(4), 479-480. doi: 10.1001/jamaophthalmol.2014.5515

Li, S. L., Jost, R. M., Morale, S. E., Stager, D. R., Dao, L., Stager, D., & Birch, E. E. (2014). A binocular iPad treatment for amblyopic children. *Eye, 28*(10), 1246-1253. doi: 10.1038/eye.2014.165

Li, S. L., Reynaud, A., Hess, R. F., Wang, Y. Z., Jost, R. M., Morale, S. E., . . . Birch, E. E. (2015). Dichoptic movie viewing treats childhood amblyopia. *J AAPOS, 19*(5), 401-405. doi: 10.1016/j.jaapos.2015.08.003

Li, T., & Shotton, K. (2009). Conventional occlusion versus pharmacologic penalization for amblyopia. *Cochrane Database Syst Rev*(4), Cd006460. doi: 10.1002/14651858.CD006460.pub2

Li, X., Dumoulin, S. O., Mansouri, B., & Hess, R. F. (2007). Cortical deficits in human amblyopia: their regional distribution and their relationship to the contrast detection deficit. *Invest Ophthalmol Vis Sci, 48*(4), 1575-1591. doi: 10.1167/iovs.06-1021

Li, X., Mullen, K. T., Thompson, B., & Hess, R. F. (2011). Effective connectivity anomalies in human amblyopia. *Neuroimage, 54*(1), 505-516. doi: 10.1016/j.neuroimage.2010.07.053

Li, Y., Fitzpatrick, D., & White, L. E. (2006). The development of direction selectivity in ferret visual cortex requires early visual experience. *Nature neuroscience, 9*(5), 676-681.

Liang, M., Xie, B., Yang, H., Yin, X., Wang, H., Yu, L., . . . Wang, J. (2017). Altered interhemispheric functional connectivity in patients with anisometropic and strabismic amblyopia: a resting-state fMRI study. *Neuroradiology, 59*(5), 517-524. doi: 10.1007/s00234-017-1824-0

Liu, X.-Y., & Zhang, J.-Y. (2017). Dichoptic training in adults with amblyopia: Additional stereoacuity gains over monocular training. *Vision Res*. doi: https://doi.org/10.1016/j.visres.2017.07.002

Longmuir, S., Pfeifer, W., Scott, W., & Olson, R. (2012). Effect of occlusion amblyopia after prescribed full-time occlusion on long-term visual acuity outcomes. *J Pediatr Ophthalmol Strabismus, 50*(2), 94-101.

Loudon, S. E., Fronius, M., Looman, C. W., Awan, M., Simonsz, B., van der Maas, P. J., & Simonsz, H. J. (2006). Predictors and a remedy for noncompliance with amblyopia therapy in children measured with the occlusion dose monitor. *Invest Ophthalmol Vis Sci, 47*(10), 4393-4400. doi: 10.1167/iovs.05-1428

Loudon, S. E., Passchier, J., Chaker, L., de Vos, S., Fronius, M., Harrad, R. A., . . . Simonsz, H. J. (2009). Psychological causes of non-compliance with electronically monitored occlusion therapy for amblyopia. *Br J Ophthalmol, 93*(11), 1499-1503. doi: 10.1136/bjo.2008.149815

Loudon, S. E., & Simonsz, H. J. (2005). The history of the treatment of amblyopia. *Strabismus, 13*(2), 93-106. doi: 10.1080/09273970590949818

Lunghi, C., Morrone, M. C., Secci, J., & Caputo, R. (2016). Binocular Rivalry Measured 2 Hours After Occlusion Therapy Predicts the Recovery Rate of the Amblyopic Eye in Anisometropic Children. *Invest Ophthalmol Vis Sci, 57*(4), 1537-1546. doi: 10.1167/iovs.15-18419

Mansouri, B., Thompson, B., & Hess, R. F. (2008). Measurement of suprathreshold binocular interactions in amblyopia. *Vision Res, 48*(28), 2775-2784. doi: 10.1016/j.visres.2008.09.002

Mayer, D. L., Beiser, A. S., Warner, A. F., Pratt, E. M., Raye, K. N., & Lang, J. M. (1995). Monocular acuity norms for the Teller Acuity Cards between ages one month and four years. *Invest Ophthalmol Vis Sci, 36*(3), 671-685.

McCormick, A., Bhola, R., Brown, L., Squirrel, D., Giles, J., & Pepper, I. (2002). Quantifying relative afferent pupillary defects using a Sbisa bar. *Br J Ophthalmol, 86*(9), 985-987.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods, 1*(1), 30.

McKee, S. P., Levi, D. M., & Movshon, J. A. (2003). The pattern of visual deficit in amblyopia. *Journal of Vision, 3*(5), 380-405. doi: 10:1167/3.5.5

MEPEDS. (2008). Multi-Ethnic Pediatric Eye Disease Study: Prevalence of amblyopia and strabismus in African American and Hispanic children ages 6 to 72 months the multi-ethnic pediatric eye disease study. *Ophthalmology, 115*(7), 1229-1236.e1221. doi: 10.1016/j.ophtha.2007.08.001

Miles, W. R. (1930). Ocular Dominance in Human Adults. *The Journal of General Psychology, 3*(3), 412-430. doi: 10.1080/00221309.1930.9918218

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends Neurosci, 6*, 414-417.

Mitchell, D. E. (2008). A special role for binocular visual input during development and as a component of occlusion therapy for treatment of amblyopia. *Restor Neurol Neurosci, 26*(4-5), 425-434.

Mitchell, D. E., & Duffy, K. R. (2014). The case from animal studies for balanced binocular treatment strategies for human amblyopia. *Ophthalmic Physiol Opt, 34*(2), 129-145. doi: 10.1111/opo.12122

Moseley, M. J., Fielder, A. R., & Stewart, C. E. (2009). The optical treatment of amblyopia. *Optom Vis Sci, 86*(6), 629-633. doi: 10.1097/OPX.0b013e3181a7b3e5

Moseley, M. J., Wallace, M. P., Stephens, D. A., Fielder, A. R., Smith, L. C., & Stewart, C. E. (2015). Personalized versus standardized dosing strategies for the treatment of childhood amblyopia: study protocol for a randomized controlled trial. *Trials, 16*, 189. doi: 10.1186/s13063-015-0711-4

Movshon, J. A., Eggers, H. M., Gizzi, M. S., Hendrickson, A. E., Kiorpes, L., & Boothe, R. G. (1987). Effects of early unilateral blur on the macaque's visual system. III. Physiological observations. *J Neurosci, 7*(5), 1340-1351.

Narasimhan, S., Harrison, E. R., & Giaschi, D. E. (2012). Quantitative measurement of interocular suppression in children with amblyopia. *Vision Res, 66*, 1-10. doi: 10.1016/j.visres.2012.06.007

Newsham, D., & O'Connor, A. R. (2016). Assessment of the Density of Suppression to Identify Risk of Intractable Diplopia in the United Kingdom. *Strabismus, 24*(2), 45-50. doi: 10.3109/09273972.2016.1170048

Niechwiej-Szwedo, E., Goltz, H. C., Chandrakumar, M., Hirji, Z., Crawford, J. D., & Wong, A. M. (2011). Effects of anisometropic amblyopia on visuomotor behavior, part 2: visually guided reaching. *Invest Ophthalmol Vis Sci, 52*(2), 795-803. doi: 10.1167/iovs.10-6092

Niechwiej-Szwedo, E., Goltz, H. C., Chandrakumar, M., & Wong, A. M. (2014). Effects of strabismic amblyopia on visuomotor behavior: part II. Visually guided reaching. *Invest Ophthalmol Vis Sci, 55*(6), 3857-3865. doi: 10.1167/iovs.14-14543

Norcia, A. M., & Tyler, C. W. (1985). Spatial frequency sweep VEP: visual acuity during the first year of life. *Vision Res, 25*(10), 1399-1408.

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

O'Connor, A. R., Birch, E. E., Anderson, S., & Draper, H. (2010). The functional significance of stereopsis. *Invest Ophthalmol Vis Sci, 51*(4), 2019-2023. doi: 10.1167/iovs.09-4434

O'Shea, R. P., Sims, A. J. H., & Govan, D. G. (1997). The effect of spatial frequency and field size on the spread of exclusive visibility in binocular rivalry. *Vision Research, 37*(2), 175-183. doi: https://doi.org/10.1016/S0042-6989(96)00113-7

O'Shea, R. P., Parker, A., La Rooy, D., & Alais, D. (2009). Monocular rivalry exhibits three hallmarks of binocular rivalry: Evidence for common processes. *Vision Res, 49*(7), 671-681. doi: http://dx.doi.org/10.1016/j.visres.2009.01.020

Ohzawa, I., Deangelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science, 249*(4972), 1037-1041.

Ooi, T. L., & He, Z. J. (2001). Sensory eye dominance. *Optometry, 72*(3), 168-178.

Ooi, T. L., Su, Y. R., Natale, D. M., & He, Z. J. (2013). A push-pull treatment for strengthening the 'lazy eye' in amblyopia. *Curr Biol, 23*(8), R309-310. doi: 10.1016/j.cub.2013.03.004

Patel, V., Stuit, S., & Blake, R. (2015). Individual differences in the temporal dynamics of binocular rivalry and stimulus rivalry. *Psychonomic bulletin & review, 22*(2), 476-482.

Pediatric Disease Investigator Group. (2003). A comparison of atropine and patching treatments for moderate amblyopia by patient age, cause of amblyopia, depth of amblyopia, and other factors. *Ophthalmology, 110*(8), 1632-1637; discussion 1637-1638.

Pediatric Eye Disease Investigator Group. (2004). A randomized trial of atropine regimens for treatment of moderate amblyopia in children. *Ophthalmology, 111*(11), 2076-2085. e2074.

Pelli, D. G., & Bex, P. (2013). Measuring contrast sensitivity. *Vision Res, 90*, 10-14. doi: https://doi.org/10.1016/j.visres.2013.04.015

Piano, M., & Newsham, D. (2015). A pilot study examining density of suppression measurement in strabismus. *Strabismus, 23*(1), 14-21. doi: 10.3109/09273972.2014.1002621

Piano, M., O'connor, A. R., & Stephenson, M. G. (2015). What does that mean? A review of some aspects of orthoptic terminology. *British and Irish Orthoptic Journal, 7*, 8-13.

Piano, M. E. F., Bex, P. J., & Simmers, A. J. (2015). Perceptual Visual Distortions in Adult Amblyopia and Their Relationship to Clinical FeaturesPerceptual Visual Distortions in Adult Amblyopia. *Invest Ophthalmol Vis Sci, 56*(9), 5533-5542. doi: 10.1167/iovs.15-17071

Poggio, G. F., & Poggio, T. (1984). The analysis of stereopsis. *Annu Rev Neurosci, 7*(1), 379-412.

Pointer, J. S. (2007). The absence of lateral congruency between sighting dominance and the eye with better visual acuity. *Ophthalmic and Physiological Optics, 27*(1), 106-110.

Pointer, J. S. (2012). Sighting versus sensory ocular dominance. *Journal of Optometry, 5*(2), 52-55. doi: 10.1016/j.optom.2012.03.001

Polat, U., Ma-Naim, T., Belkin, M., & Sagi, D. (2004). Improving vision in adult amblyopia by perceptual learning. *Proc Natl Acad Sci U S A, 101*(17), 6692-6697.

Polat, U., & Sagi, D. (1995). *Plasticity of spatial interactions in early vision.* Paper presented at the Sante Fe Institute Studies in the Sciences of Complexity-Proceedings.

Powell, C., & Hatt, S. R. (2009). Vision screening for amblyopia in childhood. *Cochrane Database of Systematic Reviews*(3). doi: DOI: 10.1002/14651858.CD005020.pub3

Pradeep, A., Proudlock, F. A., Awan, M., Bush, G., Collier, J., & Gottlob, I. (2014). An educational intervention to improve adherence to high-dosage patching regimen for amblyopia: a randomised controlled trial. *Br J Ophthalmol, 98*(7), 865-870. doi: 10.1136/bjophthalmol-2013-304187

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data., from http://www.palamedestoolbox.org

Pugh, M. (1958). Visual distortion in amblyopia. *Br J Ophthalmol, 42*(8), 449-460.

Purves, D., Augustine, G. J., & Fitzpatrick, D. (2001). *Functional specialization of the Rod and Cone Systems* (Neuroscience Ed. 2nd ed.). Sunderland (MA): Sinauer Associates.

Repka, M. X., Beck, R. W., Holmes, J. M., Birch, E. E., Chandler, D. L., Cotter, S. A., . . . Scheiman, M. M. (2003). A randomized trial of patching regimens for treatment of moderate amblyopia in children. *Arch Ophthalmol, 121*(5), 603-611. doi: 10.1001/archopht.121.5.603

Repka, M. X., Kraker, R. T., Dean, T. W., Beck, R. W., Siatkowski, R. M., Holmes, J. M., . . . Wallace, D. K. (2015). A randomized trial of levodopa as treatment for residual amblyopia in older children. *Ophthalmology, 122*(5), 874-881. doi: 10.1016/j.ophtha.2015.01.002

Repka, M. X., Kraker, R. T., Holmes, J. M., Summers, A. I., Glaser, S. R., Barnhardt, C. N., & Tien, D. R. (2014). Atropine vs patching for treatment of moderate amblyopia: follow-up at 15 years of age of a randomized clinical trial. *JAMA Ophthalmol, 132*(7), 799-805. doi: 10.1001/jamaophthalmol.2014.392

Repka, M. X., Wallace, D. K., Beck, R. W., Kraker, R. T., Birch, E. E., Cotter, S. A., . . . Weakley, D. R. (2005). Two-year follow-up of a 6-month randomized trial of atropine vs patching for treatment of moderate amblyopia in children. *Arch Ophthalmol, 123*(2), 149-157.

Rice, M. L., Leske, D. A., Smestad, C. E., & Holmes, J. M. (2008). Results of Ocular Dominance Testing Depend on Assessment Method. *J AAPOS, 12*(4), 365-369. doi: 10.1016/j.jaapos.2008.01.017

Rislove, E. M., Hall, E. C., Stavros, K. A., & Kiorpes, L. (2010). Scale-dependent loss of global form perception in strabismic amblyopia. *J Vis, 10*(12), 25. doi: 10.1167/10.12.25

Rucker, J. C., & Tomsak, R. L. (2005). Binocular diplopia. A practical approach. *Neurologist, 11*(2), 98-110. doi: 10.1097/01.nrl.0000156318.80903.b1

Rutstein, R. P., Quinn, G. E., Lazar, E. L., Beck, R. W., Bonsall, D. J., Cotter, S. A., . . . Suh, D. W. (2010). A randomized trial comparing Bangerter filters and patching for the treatment of moderate amblyopia in children. *Ophthalmology, 117*(5), 998-1004.e1006. doi: 10.1016/j.ophtha.2009.10.014

Scheiman, M. M., Hertle, R. W., Beck, R. W., Edwards, A. R., Birch, E., Cotter, S. A., . . . Tamkins, S. M. (2005). Randomized trial of treatment of amblyopia in children aged 7 to 17 years. *Arch Ophthalmol, 123*(4), 437-447. doi: 10.1001/archopht.123.4.437

Schmucker, C., Kleijnen, J., Grosselfinger, R., Riemsma, R., Antes, G., Lange, S., & Lagreze, W. (2010). Effectiveness of early in comparison to late(r) treatment in children with amblyopia or its risk factors: a systematic review. *Ophthalmic Epidemiol, 17*(1), 7-17. doi: 10.3109/09286580903312301

Scholl, B., Tan, A. Y., & Priebe, N. J. (2013). Strabismus disrupts binocular synaptic integration in primary visual cortex. *J Neurosci, 33*(43), 17108-17122. doi: 10.1523/jneurosci.1831-13.2013

Schor, C., Wood, I., & Ogawa, J. (1984). Binocular sensory fusion is limited by spatial resolution. *Vision Res, 24*(7), 661-665.

Schor, C. M., & Tyler, C. W. (1981). Spatio-temporal properties of Panum's fusional area. *Vision Res, 21*(5), 683-692.

Secen, J., Culham, J., Ho, C., & Giaschi, D. (2011). Neural correlates of the multiple-object tracking deficit in amblyopia. *Vision Res, 51*(23-24), 2517-2527. doi: 10.1016/j.visres.2011.10.011

Sengpiel, F., Baddeley, R. J., Freeman, T. C. B., Harrad, R., & Blakemore, C. (1998). Different mechanisms underlie three inhibitory phenomena in cat area 17. *Vision Res, 38*(14), 2067-2080.

Sengpiel, F., & Blakemore, C. (1996). The neural basis of suppression and amblyopia in strabismus. *Eye (Lond), 10 (Pt 2)*, 250-258. doi: 10.1038/eye.1996.54

Sengpiel, F., Jirmann, K. U., Vorobyov, V., & Eysel, U. T. (2006). Strabismic suppression is mediated by inhibitory interactions in the primary visual cortex. *Cereb Cortex, 16*(12), 1750-1758. doi: 10.1093/cercor/bhj110

Sengpiel, F., & Kind, P. C. (2002). The role of activity in development of the visual system. *Curr Biol, 12*(23), R818-826.

Sharma, V., Levi, D. M., & Klein, S. A. (2000). Undercounting features and missing features: evidence for a high-level deficit in strabismic amblyopia. *Nat Neurosci, 3*(5), 496-501.

Shooner, C., Hallum, L. E., Kumbhani, R. D., Ziemba, C. M., Garcia-Marin, V., Kelly, J. G., . . . Kiorpes, L. (2015). Population representation of visual information in areas V1 and V2 of amblyopic macaques. *Vision Res*(0).

Simmers, A. J., & Bex, P. J. (2004). The representation of global spatial structure in amblyopia. *Vision Res, 44*(5), 523-533.

Simmers, A. J., Gray, L. S., & Spowart, K. (1997). Screening for amblyopia: a comparison of paediatric letter tests. *British Journal of Ophthalmology, 81*(6), 465-469.

Simmers, A. J., Ledgeway, T., Hess, R. F., & McGraw, P. V. (2003). Deficits to global motion processing in human amblyopia. *Vision Res, 43*(6), 729-738.

Simons, K. (1983). Visual acuity norms in young children. *Survey of Ophthalmology, 28*(2), 84-92. doi: 10.1016/0039-6257(83)90076-0

Sincich, L. C., Jocson, C. M., & Horton, J. C. (2012). Neuronal projections from V1 to V2 in amblyopia. *J Neurosci, 32*(8), 2648-2656. doi: 10.1523/jneurosci.4799-11.2012

Sireteanu, R., Thiel, A., Fikus, S., & Iftime, A. (2008). Patterns of spatial distortions in human amblyopia are invariant to stimulus duration and instruction modality. *Vision Res, 48*(9), 1150-1163. doi: 10.1016/j.visres.2008.01.028

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. *J Vis, 14*(5), 3. doi: 10.1167/14.5.3

Spiegel, D. P., Baldwin, A. S., & Hess, R. F. (2016). The Relationship Between Fusion, Suppression, and Diplopia in Normal and Amblyopic Vision. *Invest Ophthalmol Vis Sci, 57*(13), 5810-5817. doi: 10.1167/iovs.16-20438

Spiegel, D. P., Byblow, W. D., Hess, R. F., & Thompson, B. (2013). Anodal Transcranial Direct Current Stimulation Transiently Improves Contrast Sensitivity and Normalizes Visual Cortex Activation in Individuals With Amblyopia. *Neurorehabil Neural Repair*. doi: 10.1177/1545968313491006

Spiegel, D. P., Li, J., Hess, R. F., Byblow, W. D., Deng, D., Yu, M., & Thompson, B. (2013). Transcranial Direct Current Stimulation Enhances Recovery of Stereopsis in Adults With Amblyopia. *Neurotherapeutics*. doi: 10.1007/s13311-013-0200-y

Spierer, A., Raz, J., Benezra, O., Herzog, R., Cohen, E., Karshai, I., & Benezra, D. (2010). Treating amblyopia with liquid crystal glasses: a pilot study. *Invest Ophthalmol Vis Sci, 51*(7), 3395-3398. doi: 10.1167/iovs.09-4568

Stewart, C. E., Fielder, A. R., Stephens, D. A., & Moseley, M. J. (2005). Treatment of unilateral amblyopia: factors influencing visual outcome. *Invest Ophthalmol Vis Sci, 46*(9), 3152-3160. doi: 10.1167/iovs.05-0357

Stewart, C. E., Moseley, M. J., & Fielder, A. R. (2011). Amblyopia therapy: an update. *Strabismus, 19*(3), 91-98. doi: 10.3109/09273972.2011.600421

Stewart, C. E., Moseley, M. J., Fielder, A. R., & Stephens, D. A. (2004). Refractive adaptation in amblyopia: quantification of effect and implications for practice. *British Journal of Ophthalmology, 88*(12), 1552.

Stewart, C. E., Moseley, M. J., Georgiou, P., & Fielder, A. R. (2017). Occlusion dose monitoring in amblyopia therapy: status, insights, and future directions. *Journal of American Association for Pediatric Ophthalmology and Strabismus*. doi: https://doi.org/10.1016/j.jaapos.2017.06.018

Stewart, C. E., Moseley, M. J., Stephens, D. A., & Fielder, A. R. (2004). Treatment dose-response in amblyopia therapy: the Monitored Occlusion Treatment of Amblyopia Study (MOTAS). *Invest Ophthalmol Vis Sci, 45*(9), 3048-3054. doi: 10.1167/iovs.04-0250

Stewart, C. E., Shah, S., Wren, S., & Roberts, C. J. (2016). Paediatric Eye Services: How Much of the Workload Is Amblyopia-Related? *Strabismus, 24*(3), 109-112. doi: 10.1080/09273972.2016.1205102

Stewart, C. E., Stephens, D. A., Fielder, A. R., & Moseley, M. J. (2007). Objectively monitored patching regimens for treatment of amblyopia: randomised trial. *BMJ, 335*(7622), 707. doi: 10.1136/bmj.39301.460150.55

Stewart, C. E., Wallace, M. P., Stephens, D. A., Fielder, A. R., & Moseley, M. J. (2013). The effect of amblyopia treatment on stereoacuity. *Journal of American Association for Pediatric Ophthalmology and Strabismus, 17*(2), 166-173.

Suttle, C., Alexander, J., Liu, M., Ng, S., Poon, J., & Tran, T. (2009). Sensory ocular dominance based on resolution acuity, contrast sensitivity and alignment sensitivity. *Clinical and Experimental Optometry, 92*(1), 2-8.

Suttle, C. M. (2010). Active treatments for amblyopia: a review of the methods and evidence base. *Clin Exp Optom, 93*(5), 287-299. doi: 10.1111/j.1444-0938.2010.00486.x

Tailor, V., Balduzzi, S., Hull, S., Rahi, J., Schmucker, C., Virgili, G., & Dahlmann-Noor, A. (2014). Tests for detecting strabismus in children age 1 to 6 years in the community. *Cochrane Database of Systematic Reviews*(7). doi: 10.1002/14651858.CD011221

Tailor, V., Bossi, M., Bunce, C., Greenwood, J. A., & Dahlmann-Noor, A. (2014). Binocular versus standard occlusion or blurring treatment for unilateral amblyopia in children aged three to eight years. *Cochrane Database of Systematic Reviews*(11). doi: 10.1002/14651858.CD011347

Tailor, V., Bossi, M., Greenwood, J. A., & Dahlmann-Noor, A. (2016). Childhood amblyopia: current management and new trends. *Br Med Bull, 119*(1), 75-86. doi: 10.1093/bmb/ldw030

Tailor, V. K., Schwarzkopf, D. S., & Dahlmann-Noor, A. H. (2016). Neuroplasticity and amblyopia: vision at the balance point. *Curr Opin Neurol*. doi: 10.1097/wco.0000000000000413

Takesian, A. E., & Hensch, T. K. (2013). Balancing plasticity/stability across brain development. *Prog Brain Res, 207*, 3-34. doi: 10.1016/b978-0-444-63327-9.00001-1

Tanaka, S., Tani, T., Ribot, J., O'Hashi, K., & Imamura, K. (2009). A Postnatal Critical Period for Orientation Plasticity in the Cat Visual Cortex. *PLoS One, 4*(4), e5380. doi: 10.1371/journal.pone.0005380

Tao, X., Zhang, B., Shen, G., Wensveen, J., Smith, E. L., 3rd, Nishimoto, S., . . . Chino, Y. M. (2014). Early monocular defocus disrupts the normal development of receptive-field structure in V2 neurons of macaque monkeys. *J Neurosci, 34*(41), 13840-13854. doi: 10.1523/jneurosci.1992-14.2014

Taylor, K., & Elliott, S. (2014). Interventions for strabismic amblyopia. *Cochrane Database Syst Rev, 7*, Cd006461. doi: 10.1002/14651858.CD006461.pub4

Teller, D. Y. (1979). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior and Development, 2*, 135-153.

Teller, D. Y. (1981). The development of visual acuity in human and monkey infants. *Trends Neurosci, 4*, 21-24.

Teller, D. Y. (1997). First glances: the vision of infants. the Friedenwald lecture. *Invest Ophthalmol Vis Sci, 38*(11), 2183-2203.

Thayaparan, K., Crossland, M. D., & Rubin, G. S. (2007). Clinical assessment of two new contrast sensitivity charts. *Br J Ophthalmol, 91*(6), 749-752. doi: 10.1136/bjo.2006.109280

The Royal College of Ophthalmologists. (2012). *Guidelines for the Management of Strabismus in Childhood*. London: Retrieved from www.rcophth.ac.uk.

Thompson, B., Mansouri, B., Koski, L., & Hess, R. F. (2008). Brain plasticity in the adult: modulation of function in amblyopia with rTMS. *Curr Biol, 18*(14), 1067-1071. doi: 10.1016/j.cub.2008.06.052

Thompson, B., Troje, N. F., Hansen, B. C., & Hess, R. F. (2008). Amblyopic perception of biological motion. *J Vis, 8*(4), 22.21-14. doi: 10.1167/8.4.22

Thompson, J. R., Woodruff, G., Hiscox, F. A., Strong, N., & Minshull, C. (1991). The incidence and prevalence of amblyopia detected in childhood. *Public Health, 105*(6), 455-462. doi: http://dx.doi.org/10.1016/S0033-3506(05)80616-X

To, L., Thompson, B., Blum, J. R., Maehara, G., Hess, R. F., & Cooperstock, J. R. (2011). A game platform for treatment of amblyopia. *IEEE Trans Neural Syst Rehabil Eng, 19*(3), 280-289. doi: 10.1109/tnsre.2011.2115255

Tsirlin, I., Colpa, L., Goltz, H. C., & Wong, A. M. (2015). Behavioral Training as New Treatment for Adult Amblyopia: A Meta-Analysis and Systematic Review. *Invest Ophthalmol Vis Sci, 56*(6), 4061-4075. doi: 10.1167/iovs.15-16583

Tytla, M. E., & Labow-Daily, L. S. (1981). Evaluation of the CAM treatment for amblyopia: a controlled study. *Invest Ophthalmol Vis Sci, 20*(3), 400-406.

Tytla, M. E., Lewis, T. L., Maurer, D., & Brent, H. P. (1993). Stereopsis after congenital cataract. *Invest Ophthalmol Vis Sci, 34*(5), 1767-1773.

van Leeuwen, R., Eijkemans, M. J., Vingerling, J. R., Hofman, A., de Jong, P. T., & Simonsz, H. J. (2007). Risk of bilateral visual impairment in individuals with amblyopia: the Rotterdam study. *Br J Ophthalmol, 91*(11), 1450-1451. doi: 10.1136/bjo.2006.113670

Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The Case for Using the Repeatability Coefficient When Calculating Test–Retest Reliability. *PLoS One, 8*(9), e73990. doi: 10.1371/journal.pone.0073990

Vedamurthy, I., Nahum, M., Bavelier, D., & Levi, D. M. (2015). Mechanisms of recovery of visual function in adult amblyopia through a tailored action video game. *Sci Rep, 5*, 8482. doi: 10.1038/srep08482

Vedamurthy, I., Nahum, M., Huang, S. J., Zheng, F., Bayliss, J., Bavelier, D., & Levi, D. M. (2015). A dichoptic custom-made action video game as a treatment for adult amblyopia. *Vision Res, 114*, 173-187. doi: 10.1016/j.visres.2015.04.008

Vojnikovic, B., & Tamajo, E. (2013). Horopters--definition and construction. *Coll Antropol, 37 Suppl 1*, 9-12.

Waddingham, P. E., Butler, T. K. H., Cobb, S. V., Moody, A. D. R., Comaish, I. F., Haworth, S. M., . . . Griffiths, G. D. (2005). Preliminary results from the use of the novel Interactive Binocular Treatment (I-BiT™) system, in the treatment of strabismic and anisometropic amblyopia. *Eye, 20*(3), 375-378. doi: 10.1038/sj.eye.6701883

Wallace, D. K., Lazar, E. L., Melia, M., Birch, E. E., Holmes, J. M., Hopkins, K. B., . . . Weise, K. K. (2011). Stereoacuity in children with anisometropic amblyopia. *J AAPOS, 15*(5), 455-461. doi: 10.1016/j.jaapos.2011.06.007

Wallace, D. K., Repka, M. X., Lee, K. A., Melia, M., Christiansen, S. P., Morse, C. L., & Sprunger, D. T. (2018). Amblyopia Preferred Practice Pattern®. *Ophthalmology, 125*(1), P105-P142. doi: https://doi.org/10.1016/j.ophtha.2017.10.008

Wallace, M. P., Stewart, C. E., Moseley, M. J., Stephens, D. A., & Fielder, A. R. (2013). Compliance with occlusion therapy for childhood amblyopia. *Invest Ophthalmol Vis Sci, 54*(9), 6158-6166. doi: 10.1167/iovs.13-11861

Wang, B., Sarnaik, R., & Cang, J. (2010). Critical Period Plasticity Matches Binocular Orientation Preference in the Visual Cortex. *Neuron, 65*(2), 246-256.

Wang, J. (2015). Compliance and patching and atropine amblyopia treatments. *Vision Res, 114*, 31-40. doi: http://dx.doi.org/10.1016/j.visres.2015.02.012

Wang, J., Neely, D. E., Galli, J., Schliesser, J., Graves, A., Damarjian, T. G., . . . Plager, D. A. (2016). A pilot randomized clinical trial of intermittent occlusion therapy liquid crystal glasses versus traditional patching for treatment of moderate unilateral amblyopia. *J AAPOS, 20*(4), 326-331. doi: 10.1016/j.jaapos.2016.05.014

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics, 33*(2), 113-120.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys, 33*(2), 113-120.

Weakley, D. R. (2001). The association between Nonstrabismic Anisometropia, Amblyopia, and Subnormal Binocularity. *Ophthalmology, 108*, 163-171.

Westheimer, G. (2013). Clinical evaluation of stereopsis. *Vision Res, 90*, 38-42. doi: 10.1016/j.visres.2012.10.005

Wheatstone, C. (1838). Contributions to the Physiology of Vision.--Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision. *Philosophical transactions of the Royal Society of London, 128*, 371-394.

White, L. E., Coppola, D. M., & Fitzpatrick, D. (2001). The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex. *Nature, 411*(6841), 1049-1052.

White, L. E., & Fitzpatrick, D. (2007). Vision and cortical map development. *Neuron, 56*(2), 327-338. doi: 10.1016/j.neuron.2007.10.011

Wiecek, E., Lashkari, K., Dakin, S. C., & Bex, P. (2015). Metamorphopsia and interocular suppression in monocular and binocular maculopathy. *Acta Ophthalmol, 93*(4), e318-320. doi: 10.1111/aos.12559

Wiesel, T. N. (1982). Postnatal development of the visual cortex and the influence of environment. *Nature, 299*(5884), 583-591.

Wiesel, T. N., & Hubel, D. H. (1963). Single cell responses in striate cortex of kittens deprived of vision in one eye. *J Neurophysiol, 26*, 1003-1017.

Wiesel, T. N., & Hubel, D. H. (1965a). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J Neurophysiol, 28*(6), 1029-1040.

Wiesel, T. N., & Hubel, D. H. (1965b). Extent of recovery from the effects of visual deprivation in kittens. *J Neurophysiol, 28*(6), 1060-1072.

Wong, A. M. F. (2012). New concepts concerning the neural mechanism of amblyopia and their clinical implications. *Can J Ophthalmol, 47*(5), 399-409. doi: 10.1016

Xi, J., Jia, W. L., Feng, L. X., Lu, Z. L., & Huang, C. B. (2014). Perceptual learning improves stereoacuity in amblyopia. *Invest Ophthalmol Vis Sci, 55*(4), 2384-2391. doi: 10.1167/iovs.13-12627

Xu, J. P., He, Z. J., & Ooi, T. L. (2011). A binocular perimetry study of the causes and implications of sensory eye dominance. *Vision Res, 51*(23-24), 2386-2397. doi: 10.1016/j.visres.2011.09.012

Yang, E., Blake, R., & McDonald, J. E. (2010). A New Interocular Suppression Technique for Measuring Sensory Eye Dominance. *Invest Ophthalmol Vis Sci, 51*(1), 588-593. doi: 10.1167/iovs.08-3076

Yao, Z., He, Z., Wang, Y., Lu, F., Qu, J., Zhou, J., & Hess, R. F. (2017). Absolute Not Relative Interocular Luminance Modulates Sensory Eye Dominance Plasticity in Adults. *Neuroscience, 367*, 127-133. doi: 10.1016/j.neuroscience.2017.10.029

Yoonessi, A., & Yoonessi, A. (2011). Functional Assessment of Magno, Parvo and Konio-Cellular Pathways; Current State and Future Clinical Applications. *Journal of Ophthalmic & Vision Research, 6*(2), 119-126.

Zele, A. J., Pokorny, J., Lee, D. Y., & Ireland, D. (2007). Anisometropic amblyopia: spatial contrast sensitivity deficits in inferred magnocellular and parvocellular vision. *Invest Ophthalmol Vis Sci, 48*(8), 3622-3631. doi: 10.1167/iovs.06-1207

Zhang, J. Y., Cong, L. J., Klein, S. A., Levi, D. M., & Yu, C. (2014). Perceptual learning improves adult amblyopic vision through rule-based cognitive compensation. *Invest Ophthalmol Vis Sci, 55*(4), 2020-2030. doi: 10.1167/iovs.13-13739

Zhou, D., Ni, N., Ni, A., Chen, Q., Hu, D. N., & Zhou, J. (2017). Association of Visual Acuity with Ocular Dominance in 2045 Myopic Patients. *Curr Eye Res, 42*(8), 1155-1159. doi: 10.1080/02713683.2017.1297464

Zhou, J., Clavagnier, S., & Hess, R. F. (2013). Short-term monocular deprivation strengthens the patched eye's contribution to binocular combination. *J Vis, 13*(5). doi: 10.1167/13.5.12

Zhou, J., Jia, W., Huang, C. B., & Hess, R. F. (2013). The effect of unilateral mean luminance on binocular combination in normal and amblyopic vision. *Sci Rep, 3*, 2012. doi: 10.1038/srep02012

Zhou, J., Reynaud, A., & Hess, R. F. (2014a). *Real-time modulation of perceptual eye dominance in humans* (Vol. 281).

Zhou, J., Reynaud, A., & Hess, R. F. (2014b). Real-time modulation of perceptual eye dominance in humans. *Proc Biol Sci, 281*(1795). doi: 10.1098/rspb.2014.1717

Zhou, J., Thompson, B., & Hess, R. F. (2013). A new form of rapid binocular plasticity in adult with amblyopia. *Sci Rep, 3*, 2638. doi: 10.1038/srep02638

Ziak, P., Holm, A., Halicka, J., Mojzis, P., & Pinero, D. P. (2017). Amblyopia treatment of adults with dichoptic training using the virtual reality oculus rift head mounted display: preliminary results. *BMC Ophthalmol, 17*(1), 105. doi: 10.1186/s12886-017-0501-8