



CrossMark  
click for updates

## Research

**Cite this article:** Vedamurthy I, Knill DC, Huang SJ, Yung A, Ding J, Kwon O-S, Bavelier D, Levi DM. 2016 Recovering stereo vision by squashing virtual bugs in a virtual reality environment. *Phil. Trans. R. Soc. B* **371**: 20150264.

<http://dx.doi.org/10.1098/rstb.2015.0264>

Accepted: 9 March 2016

One contribution of 15 to a theme issue  
'Vision in our three-dimensional world'.

### Subject Areas:

neuroscience

### Keywords:

stereopsis, virtual reality, perceptual learning, strabismus, amblyopia

### Author for correspondence:

Dennis M. Levi

e-mail: [dlevi@berkeley.edu](mailto:dlevi@berkeley.edu)

<sup>†</sup>Work was performed while at the University of Rochester.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2015.0264> or via <http://rstb.royalsocietypublishing.org>.

# Recovering stereo vision by squashing virtual bugs in a virtual reality environment

Indu Vedamurthy<sup>1,†</sup>, David C. Knill<sup>1</sup>, Samuel J. Huang<sup>1,†</sup>, Amanda Yung<sup>1</sup>, Jian Ding<sup>4</sup>, Oh-Sang Kwon<sup>1,3,†</sup>, Daphne Bavelier<sup>1,2</sup> and Dennis M. Levi<sup>4</sup>

<sup>1</sup>Department of Brain and Cognitive Sciences and Center for Visual Science, University of Rochester, Rochester, NY 14627-0268, USA

<sup>2</sup>Faculty of Psychology and Education Sciences, University of Geneva, CH-1211 Geneva 4, Switzerland

<sup>3</sup>School of Design and Human Engineering, UNIST, Ulsan 689-798, South Korea

<sup>4</sup>School of Optometry and Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA

DML, 0000-0002-5350-8639

Stereopsis is the rich impression of three-dimensionality, based on binocular disparity—the differences between the two retinal images of the same world. However, a substantial proportion of the population is stereo-deficient, and relies mostly on monocular cues to judge the relative depth or distance of objects in the environment. Here we trained adults who were stereo blind or stereo-deficient owing to strabismus and/or amblyopia in a natural visuomotor task—a 'bug squashing' game—in a virtual reality environment. The subjects' task was to squash a virtual dichoptic bug on a slanted surface, by hitting it with a physical cylinder they held in their hand. The perceived surface slant was determined by monocular texture and stereoscopic cues, with these cues being either consistent or in conflict, allowing us to track the relative weighting of monocular versus stereoscopic cues as training in the task progressed. Following training most participants showed greater reliance on stereoscopic cues, reduced suppression and improved stereoacuity. Importantly, the training-induced changes in relative stereo weights were significant predictors of the improvements in stereoacuity. We conclude that some adults deprived of normal binocular vision and insensitive to the disparity information can, with appropriate experience, recover access to more reliable stereoscopic information.

This article is part of the themed issue 'Vision in our three-dimensional world'.

## 1. Introduction

Stereopsis is the impression of three-dimensionality—of objects 'popping out in depth'—that most humans get when they view real-world objects with both eyes, based on binocular disparity, the differences between the two retinal images of the same world. However, a substantial proportion of the population is stereoblind or stereo-deficient. The exact proportion depends on the specific test for stereopsis and the age of the subjects, but estimates of impaired stereopsis range from  $\approx 5\%$  [1] to as high as 34% in older subjects [2]. This impairment may have a substantial impact on visuomotor tasks, difficulties in playing sports in children and locomoting safely in older adults, and may also limit career options (see ref. [3] for a recent review).

Over the past 5 years, there has been a renewed interest in restoring stereopsis in adults with strabismus since the publication of '*fixing my gaze*' [4], in which Susan Barry, a neuroscientist, recounts her recovery from strabismus (a turned eye) and her amazement as she regains stereovision, and the description by Bruce Bridgeman, a vision scientist who had been stereo-deficient all his life, of experiencing stereoscopic depth perception after viewing the three-dimensional movie Hugo [5]. However, there are a limited number of experimental studies documenting recovery of stereopsis in adults who have long been deprived of normal binocular vision. Nakatsuka *et al.* [6] reported that adult monkeys reared with

prisms had mild stereo deficiencies that improved through perceptual learning (PL) after 10 000–20 000 trials. Astle *et al.* [7] reported on two cases of humans with anisometropic amblyopia whose stereopsis improved following a learning-based course of training, which included refractive adaptation followed by monocular PL as well as stereoscopic PL. Ding & Levi [8] provided the first evidence for the recovery of stereopsis through PL in human adults long deprived of normal binocular vision owing to strabismus and/or amblyopia. They used a training paradigm that combined monocular cues that were perfectly correlated with the stereoscopic cues. Following PL (thousands of trials) with stereoscopic gratings, adults who were initially stereoblind or stereo-deficient showed substantial recovery of stereopsis. Importantly, these subjects reported that depth ‘popped out’ in real life, and they were able to enjoy three-dimensional movies for the first time, similar to the experiences of Susan Barry and Bruce Bridgeman. Their recovered stereopsis is based on perceiving depth by detecting binocular disparity, but has reduced resolution and precision. Similar improvements were recently reported in a group of anisometropic and ametropic amblyopes who were trained with anaglyphic textures with different disparities [9].

How does training improve stereopsis? There are multiple cues to depth—both binocular (retinal disparity, convergence) and monocular (motion parallax, relative size, familiar size, cast shadows, occlusion, accommodation, texture gradient, linear perspective, aerial perspective, shading, lighting and defocus blur). Stereo blind or deficient observers rely mainly on monocular cues. Ding & Levi [8] speculated that stereoblind or stereo-deficient observers could learn to associate monocular and binocular cues to depth if they were highly correlated through repeated practice. In this study, we trained the adult observers who were stereo-deficient owing to strabismus and/or amblyopia (lazy eye) in a natural visuomotor task—a ‘bug squashing’ game—in a virtual reality (VR) environment. The subjects’ task was to squash a virtual dichoptic bug on a slanted surface, by hitting it with a cylinder. The slant of the surface was determined by (i) purely stereoscopic cues (pure stereo-cue trials) or (ii) consistent monocular texture and stereoscopic cues (cue-consistent trials) or (iii) conflicting monocular texture and stereoscopic cues (cue-conflict trials). Importantly, our bug squashing training involves integrating not just multiple visual cues, but also the rich information from tactile and kinesthetic feedback. We hypothesized that training with multiple cues to depth with rich feedback might enable stereoblind or stereo-deficient observers to increase their reliance on stereoscopic cues. Following training, these observers showed increased reliance on stereoscopic cues (relative to monocular cues), reduced interocular suppression and significantly improved stereoacuity.

Our results have important implications for the recovery of visual function late in life, well outside the childhood period, until recently thought to offer the only real scope for plasticity. More broadly, they demonstrate the use of VR as a promising approach for perceptual training of all kinds.

## 2. Methods

### (a) Participants

Eleven adults (mean age 34.7 years, range 19–56 years) with long-standing abnormal binocular vision completed the training

study. Informed consent was obtained from all subjects conforming to the guidelines of the Research Subjects Review Board at the University of Rochester. Subjects were recruited mainly through referrals from local eye doctors and through print advertisements, and were paid \$10 per hour for study participation. The inclusion criteria for the experimental group were (i) impaired stereopsis associated with one or more of the following conditions—anisometropic amblyopia, strabismic amblyopia, mixed (both anisometropic and strabismic) or pure strabismus, i.e. without amblyopia, (ii) normal ocular and general health and (iii) no history of eye surgeries except for those to correct strabismic deviation. Subjects with non-comitant and/or large angle strabismus (more than 30 prism diopters) were excluded. Anisometropia was defined as greater than or equal to one-diopter difference in spherical equivalent refraction between the two eyes. Those with manifest ocular deviation (strabismus), as indicated by the cover test and no anisometropia, were classified as pure strabismics, and those having both anisometropia and strabismus were classified as mixed etiology. The clinical details of subjects are summarized in table 1. Nine adults with normal acuity and binocular vision were recruited to provide normal control stereo weight data at baseline before any training. Three of these participants were then entered in the training phase, with two completing 30 training sessions and one completing 20 training sessions. Mean results for these three normal control observers are shown in figures 3, 5 and 6–8.

Subjects completed the training study in five phases—screening, pre-testing, training, post-testing and follow-up. In the screening phase, subjects received a complete eye examination to determine whether or not they met the study inclusion criteria. Qualified subjects returned to the laboratory to complete a pre-training test battery (pre-testing) that included baseline assessment of stereopsis, suppression, vergence and visual acuity. They then underwent a VR-based bug-squashing programme (described below) for  $\approx 35$  sessions distributed over eight to 11 weeks. The pre-training test battery was re-administered a few days after training to assess any training-related changes (post-testing), and for the third time after a two month period of no intervention to assess retention effects (follow-up).

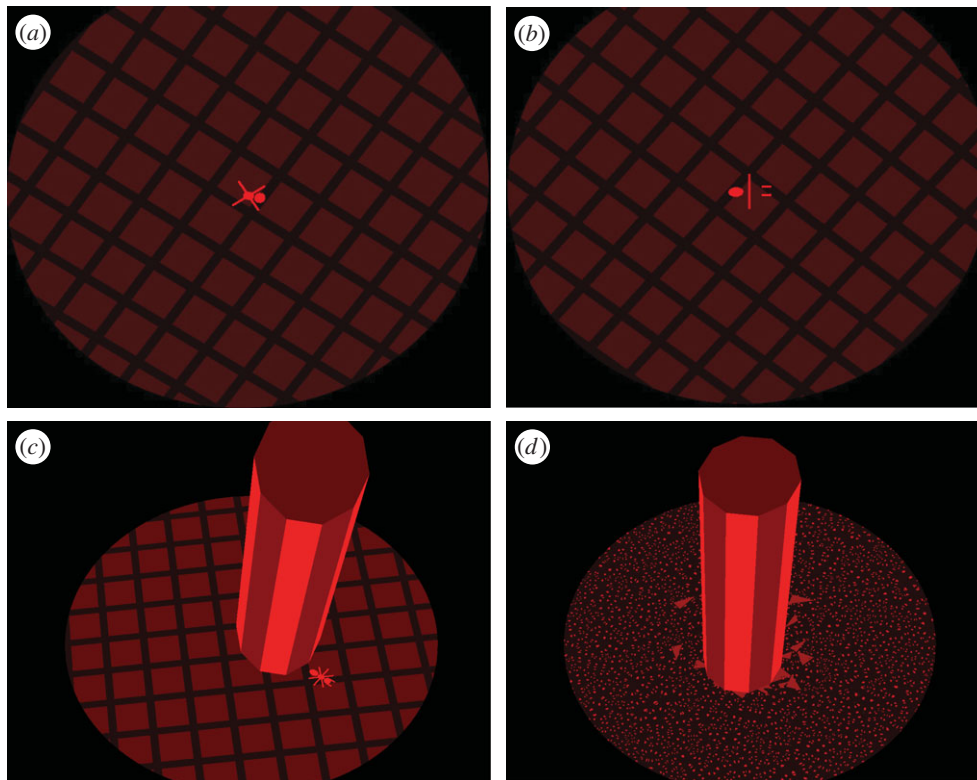
### (b) Training

The visual stimuli consisted of textured, slanted virtual discs with a central fixation target presented in a VR display (described in detail in [10]). Subjects wore Crystal Eyes shutter goggles during the training to view the stimuli (StereoGraphics Corporation, San Rafael, CA). All stimuli were drawn in red to minimize interocular cross-talk by using the relatively fast red phosphor of the monitor. A small dichoptic bug was rendered in the plane of the disc and served as the fixation target (figure 1*a,b*). In the absence of suppression, and with accurate vergence, subjects perceive a complete bug with six legs and a pair of antennae. If the non-dominant eye is suppressed, then the observer sees only the bug’s thorax, abdomen and four diagonal legs. If the dominant eye is suppressed, then the observer sees only the head and remaining legs. We used two types of textures to render the discs, so as to create stimuli with and without effective monocular cues to slant in depth—regular tiled textures, for which subjects with normal binocularity assign nearly equal weights to monocular and stereo cues (figure 1*c*), and randomly shaped dot textures (figure 1*d*) for which the texture cues are relatively uninformative [10,11].

The subjects’ task was to squash the virtual bug by hitting it with a Plexiglass cylinder measuring 6.4 cm in diameter and 12.7 cm in height and weighing 227 g. Subjects grabbed the cylinder and moved it from a starting plate positioned to the right of the virtual image (figure 2), and were required to place it as flush as possible with the target surface over the bug. A three-camera Optotrak 3020 motion-capture system (Northern Digital Inc.,

**Table 1.** Observer visual and demographic characteristics. (1) Type; A, anisometropia; M, both strabismus and anisometropia; S, pure strabismus (2) NDE, non-dominant eye (R, right; L, left); (3) ocular alignment; ortho, orthophoria; XP, exophoria; XT, exotropia; ET, esotropia; AXT, alternating exotropia; HyperT, hypertropia; HypoT, hypotropia; (4) Stereoacuity; Randot stereoacuity. F, failed ( $>400$  arcsec); note that treatment history includes any treatment beyond refractive correction with glasses or contact lenses. Age appropriate near correction was used for the various test distances. Units: visual acuity is given in logMAR units. Subjects A1, A2, M1, M4 and M5 were contact lens wearers. ETRDS VA (logMAR), early treatment diabetic retinopathy study visual acuity (logarithm of the minimum angle of resolution); plano, no refractive error.

type	age (years)/gender	NDE	refractive correction	ETDRS VA (logMAR)		ocular alignment (prims diopters)		stereo acuity (arc sec) pre/post	treatment history
				distance	near	distance	near		
A1	19/F	R	R: $-12.00/-2.00 \times 10$ L: $+0.50/-2.75 \times 5$	R: 0.2 L: 0.0	ortho	ortho	200/20	detected at age 4, no treatment	
A2	19/M	L	R: plano L: $+4.50/-1.00 \times 180$	R: $-0.1$ L: 0.16	ortho	3 XP	200/70	detected at age 11, patched at age 11 for one year	
M1	23/F	R	R: $-5.00/-2.25 \times 20$ L: $-6.50/-2.25 \times 170$	R: 0.02 L: 0.0	RXT AXT, 12	RXT AXT, 10	70/20	detected at age 2, no treatment	
M2	38/F	L	R: plano L: $+1.75/-0.50 \times 166$	R: $-0.1$ L: 0.06	12 LXT	12 LXT	$>400/>400$	detected in infancy. Two surgeries to correct strabismus	
M3	20/F	L	R: plano L: $+2.25/-0.75 \times 165$	R: $-0.1$ L: 0.08	10 LXT, 2 L HypoT	10 LXT, 2 L HypoT	200/ $>400$	detected at age 2, patched for 2 years.	
M4	30/F	R	R: $-5.25/-2.00 \times 2$ L: $-3.25/-0.50 \times 15$	R: 0.18 L: 0.14	4 RET	4 RET	200/70	detected at age 19, no treatment	
M5	23/F	R	R: $+3.50/-0.5 \times 10$ L: $+0.75$	R: 0.16 L: $-0.1$	5 RET	5 RET	$>400/70$	detected at age 13	
S1	56/F	R	R: $-1.00/-1.00 \times 90$ L: $-1.00/-1.25 \times 85$	R: 0.04 L: 0.02	RXT AXT, 22	RXT AXT, 24	200/20	detected at age $<10$ years, no treatment	
S2	55/F	R	R: $+3.00$ L: $+3.75/+0.25 \times 26$	R: 0.12 L: 0.04	3RET	3 RET	$>400/>400$	detected at age 3, patched few weeks	
S3	52/F	R	R: $-2.25/-0.75 \times 30$ L: $-2.50/-0.75 \times 170$	R: 0.24 L: $-0.06$	6 RET, 2 LHyper T	6 RET, 2 LHyper T	400/400	detected at age 2, patched for 1 year	
S4	47/F	R	R: $-4.25$ L: $-4.25$	R: 0.20 L: 0.0	6 RET	5 RET	200/200	detected at age 31, no treatment	



**Figure 1.** Examples of fixation 'bug' stimulus and texture surfaces. (a) The fixation bug's thorax, abdomen and four diagonal legs were presented to the dominant eye. (b) The head and remaining bug parts were presented to the non-dominant eye. When fused, a complete bug with six legs and a pair of antennae would be perceived. Note that figure 1a,b included here are for illustrative purposes only and were not designed as a stereo pair. (c) Tiled texture background and (d) random dot background. For tiled backgrounds rendered stereoscopically, the monocular cues to surface slant stem from distortions of the regular grid pattern mapped onto the surface, and the disparity signal to surface slant stems from the differences between right and left eyes' retinal images. Examples of trial feedback are also depicted here. The dichoptic bug crawls away (c) for an incorrect response and (d) it explodes in little bits for a correct response. (Online version in colour.)

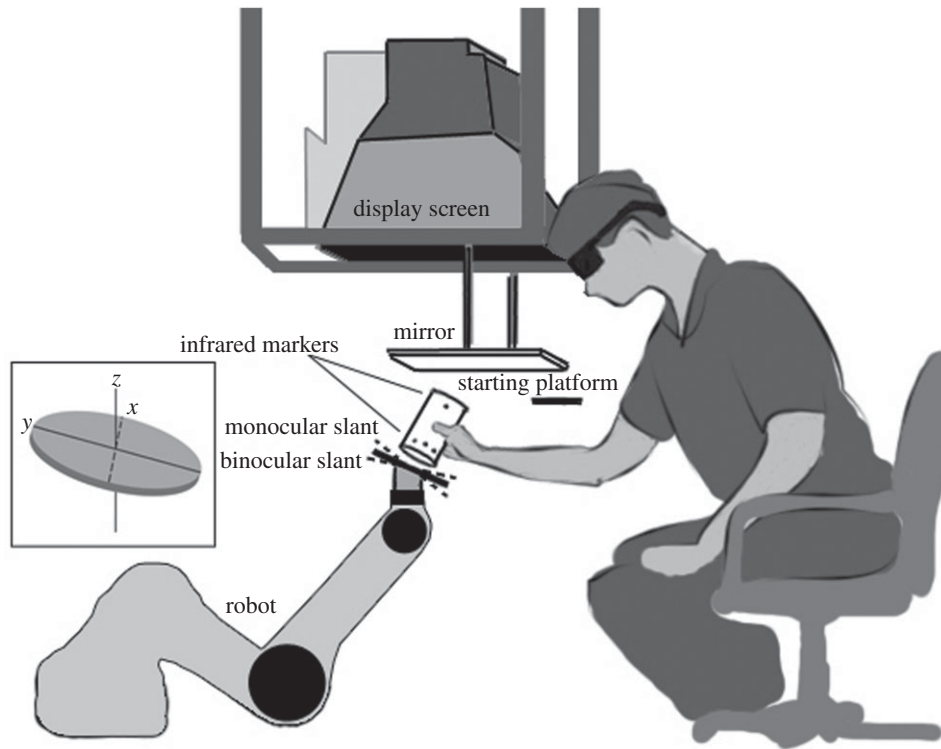
Waterloo, Canada) was used to track the four infrared markers on the cylinder so as to estimate its three-dimensional position and orientation in real-time. A virtual cylinder was rendered co-aligned with the real cylinder as it moved within the workspace. The cylinder was rendered with perspective and stereoscopic cues. As described below, because subjects viewed the stimuli through circular apertures to eliminate contextual cues provided by the monitor, the virtual cylinder was visible to the subject only when it approached the bug (typically for about approx. 250–300 ms before impact). When the cylinder was not within the subject's field of view, the only information available about the orientation of the cylinder was proprioceptive information. Every trial was programmed to end when the cylinder contacted the target plate, which provided subjects with haptic feedback about the target slant.

At the beginning of each session, subjects performed a calibration procedure to estimate the positions of their eyes relative to the monitor to ensure accurate rendering of the virtual three-dimensional space (see [10] for details). The method of adjustment was then used to equalize the perceived contrast of the bug's half-images presented to the dominant and non-dominant eyes. Measurements were repeated six times, and the mean contrast was used to display the dominant eye's nonius bug during the training task. Prisms were used to optically align the nonius bug parts for experimental subjects with strabismus. No subject reported diplopic background textures post-alignment. Contextual cues from monitor edges and other surrounding objects were removed by placing an adjustable circular aperture before each eye that restricted the viewing angle to  $11.9^\circ$ .

Participants completed 35 sessions over a period of eight to 11 weeks, with an average of three sessions per week in the training phase. The first three usable sessions (pre) and last two

sessions (post) were used to estimate changes from pre- to post-training stereo weights. Two more sessions were run two months after training to estimate changes at follow-up. In each session, target stimuli were rendered at five possible slants starting at  $20^\circ$  and ending at  $50^\circ$  away from frontoparallel in  $7.5^\circ$  steps, with one of the five slants chosen randomly between trials. Sessions were run in blocks of 60 trials, with six blocks per session totalling 360 trials. Two hundred and forty of these trials incorporated stimulus discs with tiled texture. In half of those trials, the monocular (tiled texture) and stereoscopic (disparity) cues to the surface slant were consistent (cue-consistent trials). In the other half, the slant specified by stereoscopic and texture cues differed by  $7.5^\circ$  (for example, monocular/stereoscopic slants:  $42.5/35$  or  $27.5/35$ —cue-conflict trials). The remaining 120 trials contained random dot textures that provided only stereoscopic cues (pure stereo-cue trials).

Each VR training session began only after subjects reported that the dichoptically rendered bug halves were equally visible and aligned, ensuring fusion. Thereafter, each trial commenced with participants placing the cylinder squarely on the starting plate. This signalled the robot arm to orient a real surface, co-aligned with the slant specified for that trial. For cue-consistent trials, the slant of the real surface matched the slant suggested by (consistent) monocular and stereoscopic cues. For cue-conflict trials, the robot arm oriented the real surface at the average of the slants specified by the monocular and stereoscopic cues. A slanted virtual disk with the bug was then displayed for 2 s. At the end of 2 s, the bug spun  $360^\circ$  in 250 ms (one full turn) to indicate the 'go' signal. Subjects were given 1.3 s after the go signal to move the cylinder from the starting plate and place it flush on the target surface to squash the virtual bug. Subjects were instructed to squash the bug only if both halves of the dichoptic bug were visible and aligned. The feedback strategy



**Figure 2.** Stimuli were displayed on an inverted 22-inch Mitsubishi Diamond Pro 2070SB CRT monitor, with screen resolution  $1152 \times 864$  pixels and refresh rate 120 Hz. Observers viewed the reflections of the stereoscopically presented stimuli through a mirror using Crystal Eyes shutter glasses (StereoGraphics Corporation, San Rafael, CA). An opaque black plate was placed beneath the mirror, so that subjects could see only the virtual image of the display formed below the mirror. A PUMA 260 robot arm, invisible to the subject, coaligned a circular metal (target) surface with the virtual image at a distance of 63.5 cm from the viewer's eyes. The slant of the target surface was defined as its orientation around the  $x$ -axis relative to subject's line of sight (see dashed line in the figure inset), and a  $0^\circ$  slant indicated frontoparallel. Monocular and binocular slants differed on cue-conflict trials, but subjects perceived the stimuli as a single slanted disc. Head movements were restricted with head and chin rests. Subjects' task was to move a cylinder from the starting platform and place it flush onto the real surface to squash the virtual image. The slant of the real surface was determined by the mean of the monocular and binocular slants. The starting platform was 40 cm to the right of the target surface, 20 cm closer to the subject than the target surface, and 16.5 cm above the target surface (all measured from the subject's point of view).

was identical for both cue-conflict and cue-consistent trials. If the orientation of the cylinder was within  $5^\circ$  of the orientation of the target surface, then the bug exploded. If not, the bug quickly ran across the surface and disappeared. When subjects took longer than 1.3 s, the trial was discarded, and a message was displayed instructing the participants to respond faster. The discarded trials were randomly administered later in the block. The same procedure was repeated until the entire block was completed. Likewise, if subjects moved the cylinder prior to the go signal, then the computer aborted the trial, displayed an error message and repeated the trial later in the same block. For both cue-consistent and cue-conflict trials, in order to avoid a participant being able to tell where the robot arm moved based on auditory cues, the robot arm was made to move to two different slant angles before settling on the slant specified for that trial. The two slants were randomly chosen from  $10^\circ$  to  $60^\circ$ .

After a few initial training sessions, subjects took roughly 7–8 min to complete a block. Including calibration, each session lasted 60–75 min. Video game-like scores were given for correct and incorrect responses to motivate subjects. In order to make sure that subjects were not suppressing during the task and were accurately withholding movements on trials in which they did not perceive a bug, an additional 16 trials per block were included as 'no go' trials, which contained only one half of the nonius bug (presented to the dominant eye only) for fixation. Performance on these trials was, however, not recorded and will thus not be discussed further.

Nine subjects with normal vision provided baseline measurements for this task by participating in the first three sessions of the VR task. Three of these subjects continued with the VR task, one of them for 20 sessions and the other two for 30

sessions. Although these subjects were just run to pilot the VR task, we report their data below for qualitative comparison with the stereo-deficient patients.

### (c) Analysis of virtual reality training data

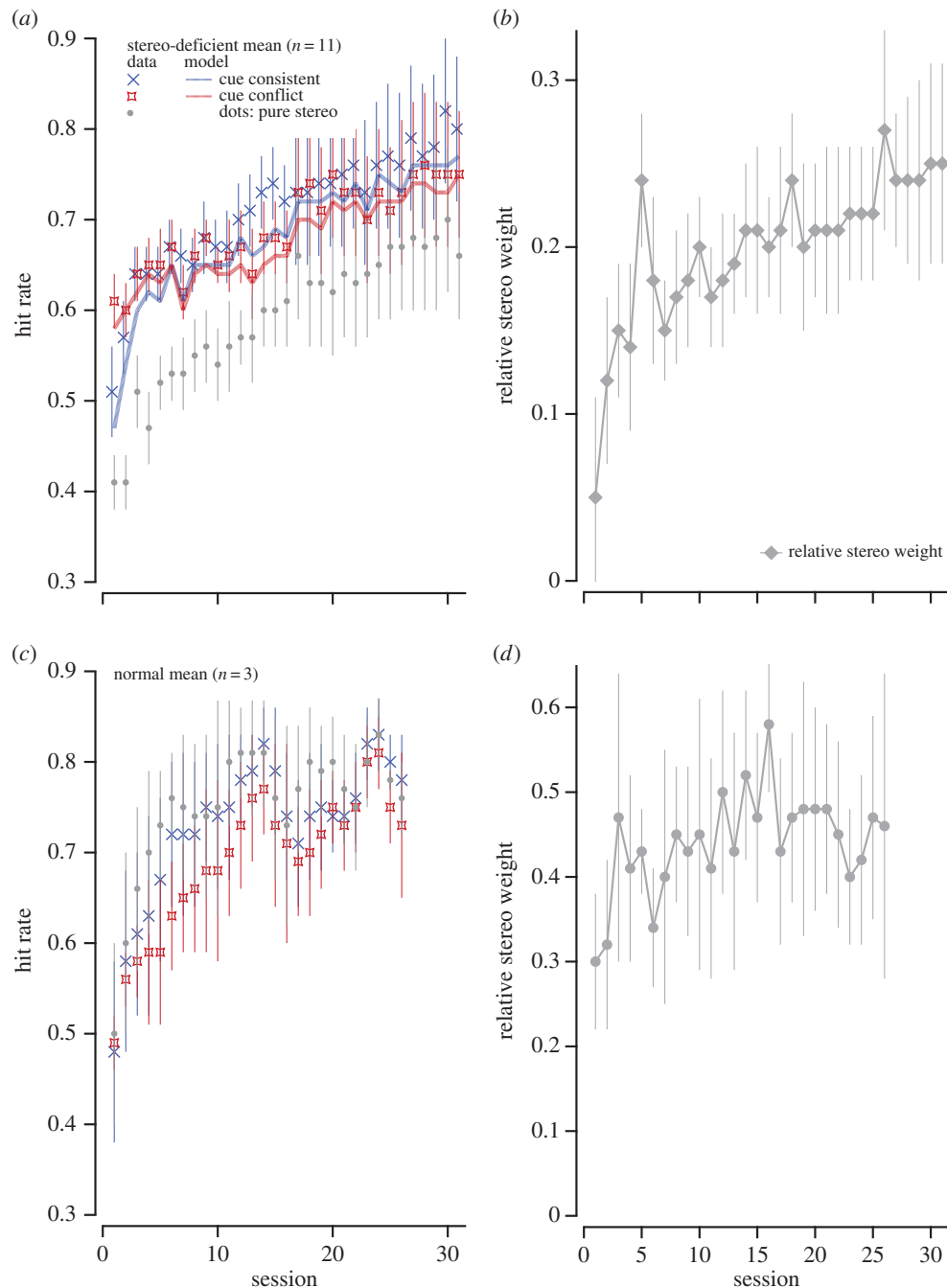
In order to measure the influence of stereoscopic cues on the surface slant estimate used by subjects to plan their hitting movements, we regressed the slant of the cylinder just prior to contact with the surface (three optotrak frames—24 msec—prior to contact) against the slant depicted by the texture on the surface and the slant depicted by stereoscopic disparities using the equation

$$s_{\text{cyl}} = w_{\text{mono}}s_{\text{mono}} + w_{\text{stereo}}s_{\text{stereo}} + k,$$

where  $s_{\text{cyl}}$  is the slant of the cylinder just prior to making contact with the surface,  $s_{\text{mono}}$  is the slant suggested by texture cues and  $s_{\text{stereo}}$  is the slant suggested by stereo cues. We normalized the weights to obtain a measure of the relative weight that subjects give to stereo to plan their movements

$$\text{Relative\_stereo\_weight} = \frac{w_{\text{stereo}}}{w_{\text{stereo}} + w_{\text{mono}}}.$$

Only trials containing regular tiled texture, and therefore, effective monocular cues, were used for the regression. Trials were further limited to those containing slants between  $27.5^\circ$  and  $42.5^\circ$  to minimize any effects on the regression that nonlinearities in the mapping between stimulus and perceived slant might have. The cue-consistent trials were pooled together with the cue-inconsistent trials to determine the regression coefficients.



**Figure 3.** Accuracy (*a,c*) and relative stereo weights (*b,d*) as training progresses. Left column: mean ‘hit rate’ for the 11 stereo-deficient (*a*) and for three stereo-normal observers (*c*). A trial was considered a hit if the orientation of the cylinder was within plus or minus  $5^\circ$  of the orientation of the target surface (the plate), which was set to be at the mean orientation of the stereoscopic and monocular cues. The lines are the model output with the parameters shown in figure 4 (see the text). Right column: the mean of the relative stereoscopic weights is plotted for the 11 stereo-deficient (*b*) and for three stereo-normal observers (*d*; see §2c of the data analysis for how these weights were derived).

## (d) Behavioural assessments

### (i) Suppression

Subjects adjusted the relative contrast of the dominant eye’s stimulus (the half bug) to match the perceived contrast of the non-dominant eye’s half bug before each training session. We use the interocular contrast ratio (ratio of non-dominant to dominant eye contrast) that appears equal as a measure of interocular suppression (a ratio of 1 = no suppression; high values = strong suppression).

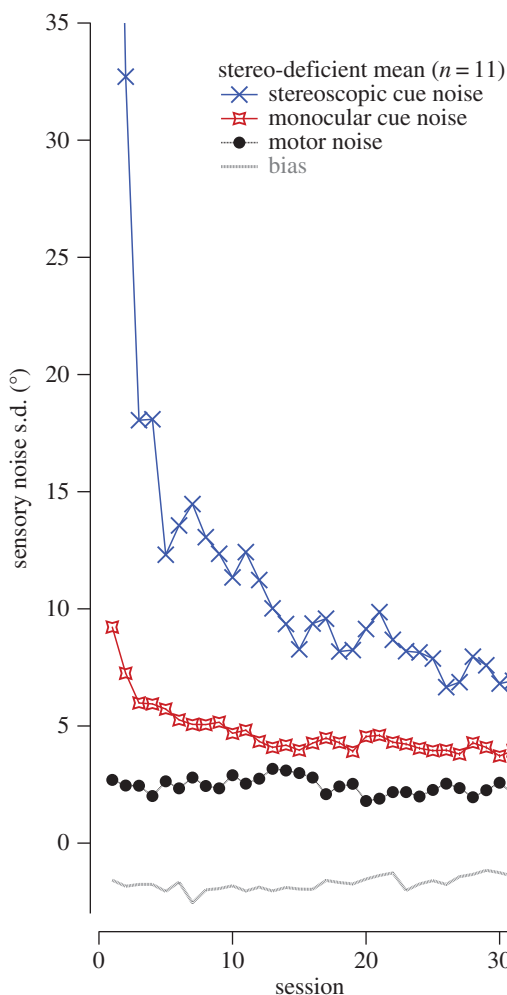
### (ii) Stereoacuity

We used a standard clinical stereo-test to evaluate changes in stereoacuity—the Randot Circles Stereotest (Stereo Optical Co., Inc.—see [12] for details). Because this test contains monocular

cues, we also included the pure disparity test (PDT using 1 cpd sine wave grating stimuli) described by Ding & Levi [8], which contains no monocular cues.

### (iii) Vergence control test

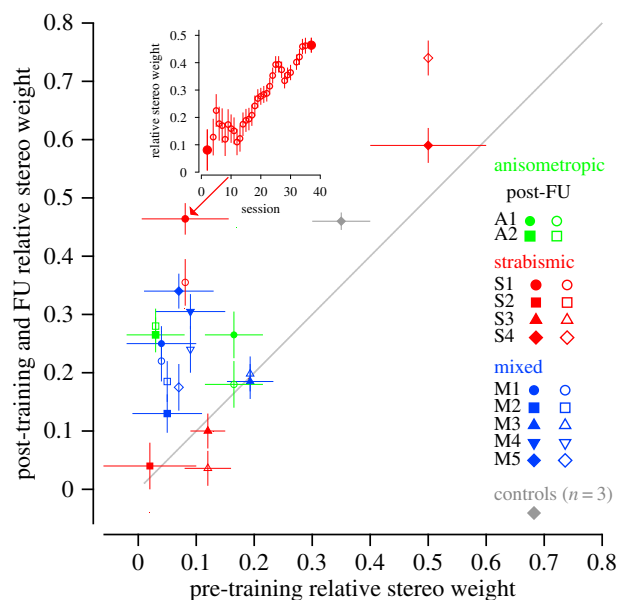
Vergence instability may negatively impact subjects’ stereo-performance. We developed a novel psychophysical test to track the effect of training on vergence accuracy. Test stimuli consisted of two thin vertical lines presented either monocularly or dichoptically using Crystal Eyes shutter goggles. These were surrounded by a fusion frame consisting of four small wedge-shaped markers with a central fixation dot shown binocularly at zero disparity. Stimuli were viewed from a test distance of 1.5 m. Phase 1 of the test required subjects to adjust the contrast of the dominant eye’s stimulus (method of adjustments) so that



**Figure 4.** Output of a simple Bayesian model fit to the mean accuracy data in the 11 stereo-deficient subjects. As participants trained in the task a significant reduction in stereoscopic and monocular cue noise is observed, with motor noise and bias remaining stable throughout training. Thus, participants' estimate of depth cues became more reliable with training.

it perceptually matched that of the non-dominant eye. Phase 2 generated a baseline measure of monocular line-alignment acuity of the dominant eye by presenting the test lines monocularly (using the contrast estimated in phase 1). Subjects were instructed to determine if the target (top) vertical line was displaced to the right or left with respect to the bottom (reference) vertical line, and respond accordingly by right or left mouse clicks. A random horizontal jitter was applied equally to both vertical lines in every trial to prevent the use of the central fixation point as a reference for target displacement. Four interleaved staircases were constructed using three-right/one-left, one-right/three-left, five-right/one-left and one-right/five-left rules. Phase 3 of the test required subjects to perform the same task as described in phase 2, but using dichoptically presented lines and a binocularly presented fusion frame. This measure of dichoptic nonius alignment is similar to that described by McKee & Levi [13]. The target line was presented to the dominant eye, and the reference line was presented to the non-dominant eye. Additionally, as a suppression check, subjects were instructed to click the mouse wheel if they only saw one line. Three consecutive 'suppression' responses warranted readjusting the line contrasts to the dominant eye, and the experiment was reset. A total of 600 trials were run.

Cumulative Gaussian psychometric functions were fit to the monocular and dichoptic test data. The vergence noise was



**Figure 5.** Post- versus pre-training (solid symbols) and follow-up (open symbols) relative stereo weights. Each coloured symbol shows the data of a single stereo-deficient observer. The grey diamond shows the pre- and post-training mean relative stereo weights of the three normal control observers who underwent training. Data above the diagonal unity line indicate increases in relative stereo weights. The inset shows the evolution of the increased relative stereo weights for strabismic subject S1 (indicated in the main figure by an arrow). Solid symbols in the inset represent the pre-to-post-training weights and the open symbols represent the pre-to-follow-up training weights.

estimated from the standard deviation parameters of the cumulative Gaussian psychometric function fits for the monocularly and dichoptically presented line stimuli,  $\sigma_{\text{vergence}} = \sqrt{\sigma_{\text{dichoptic}}^2 - \sigma_{\text{monocular}}^2}$ . The 50% point on the psychometric function reflects bias or the constant shifts in eye alignment.

### (e) Statistical analyses

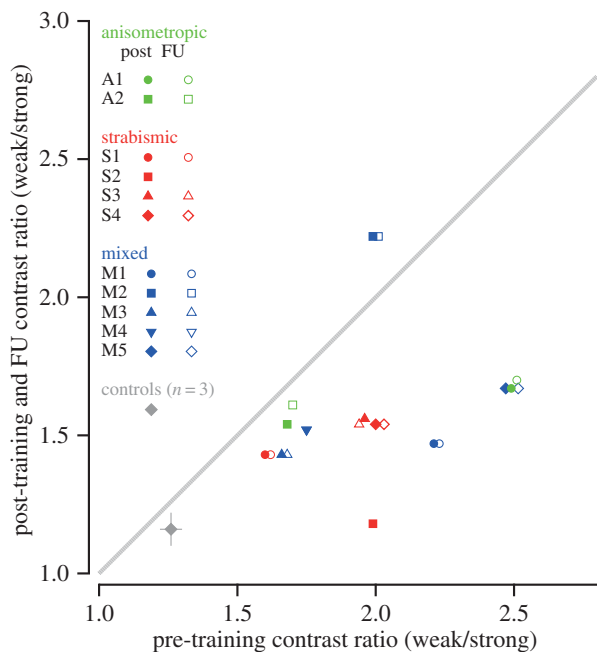
Data analyses were performed using the SPSS package. To test the effect of time, we used one-way repeated-measures ANOVA, including the three time points, pre-test, post-test and follow-up. These were followed by Bonferroni-adjusted pairwise comparisons to better identify where the specific differences lie.

Some of the measurements violated the normality assumption however; this was the case for suppression, Randot and vergence measurements. For those, the Friedman statistical test was applied. This test is the non-parametric alternative to the one-way-repeated measures ANOVA. Wilcoxon signed-rank tests were conducted to identify where the specific differences lie, with Bonferroni correction to control for inflation of type I errors. All  $p$ -values reported are two-tailed, except in pairwise tests as mentioned.

## 3. Results

### (a) Training increases the accuracy of performance

Training resulted in improved accuracy of slant judgements for both dot stimuli (slant specified by stereoscopic cues only) and for textured stimuli containing either consistent monocular and stereoscopic cues, or conflicting monocular and stereoscopic cues, for both normal and stereo-deficient observers. Figure 3 shows the mean 'hit rate' over sessions



**Figure 6.** The effect of training on suppression. Pre- versus post-training (solid symbols) and follow-up (open symbols) interocular contrast ratio (weak/strong). Each coloured symbol shows the data of a single stereo-deficient observer. The grey diamond shows the pre- and post-training mean data of the three normal control observers who underwent training. Data below the diagonal unity line indicate reduced suppression. A horizontal offset has been applied to avoid symbols (post and FU) from overlapping.

for the 11 stereo-deficient ( $a$ ) and for the three stereo-normal observers ( $c$ ). A trial was considered a hit trial if the orientation of the cylinder was within plus or minus  $5^\circ$  of the orientation of the target plate. The mean ‘hit rate’ was the average hit rate across all five slants.

In order to understand the improvement in performance, we fitted a simple Bayesian model to the mean accuracy data. The model has four parameters: stereoscopic cue noise, monocular cue noise, motor noise and a constant bias ( $k$ ). Sensory and motor noise are assumed to follow a zero mean Gaussian distribution with variances  $\sigma_{\text{stereo}}^2$ ,  $\sigma_{\text{mono}}^2$  and  $\sigma_{\text{motor}}^2$ . We assume that subjects have a Gaussian prior distribution of surface slants whose variance reflects the actual variance of stereoscopic and monocular cues presented in the task ( $\sigma_{\text{prior}} = 9.2^\circ$ ) [14], and integrate this prior knowledge with the stereoscopic and the monocular sensory signals in a statistically optimal fashion. Motor noise and bias are then added to the integrated sensory estimation. The resulting slant of the cylinder predicted by the model follows a Gaussian distribution whose mean ( $\mu_\theta$ ) and variance ( $\sigma_\theta^2$ ) can be computed as follows

$$\mu_\theta = w_{\text{sense}}(w_{\text{stereo}}\theta_{\text{stereo}} + (1 - w_{\text{stereo}})\theta_{\text{mono}}) + k,$$

where

$$w_{\text{sense}} = \frac{\sigma_{\text{prior}}^2 \sigma_{\text{stereo}}^2 + \sigma_{\text{prior}}^2 \sigma_{\text{mono}}^2}{\sigma_{\text{mono}}^2 \sigma_{\text{stereo}}^2 + \sigma_{\text{prior}}^2 \sigma_{\text{stereo}}^2 + \sigma_{\text{prior}}^2 \sigma_{\text{mono}}^2}$$

and

$$w_{\text{stereo}} = \frac{\sigma_{\text{mono}}^2}{\sigma_{\text{mono}}^2 + \sigma_{\text{stereo}}^2},$$

where  $\theta_{\text{stereo}}$  is the slant of the surface in stereoscopic cue, and  $\theta_{\text{mono}}$  is the slant of the surface in monocular cue.

$$\sigma_\theta^2 = \frac{w_{\text{sense}}^2 (\sigma_{\text{mono}}^2 \sigma_{\text{stereo}}^2)}{\sigma_{\text{mono}}^2 + \sigma_{\text{stereo}}^2} + \sigma_{\text{motor}}^2$$

As can be seen in figure 4, the stereo-deficient observers show a substantial reduction in stereoscopic cue noise over the course of training, a much smaller reduction in monocular cues noise, and essentially no change in the motor noise or bias parameters. The blue and red lines in figure 3*a* show the simulated change in hit rate based on these model parameters.

### (b) Training increases the weighting of stereo cues

The principal outcome measure, derived from the bug squashing data, is the relative stereo weight used by subjects to plan their hitting movements for stimuli containing regular textures. For these stimuli, the nine normally sighted control observers had strong relative stereo weights ( $0.39 \pm 0.05$ ) when tested under binocular conditions. Four of these subjects were retested with the dominant eye patched. Under these monocular conditions, their stereo weights were negligible ( $0.03 \pm 0.02$ ).

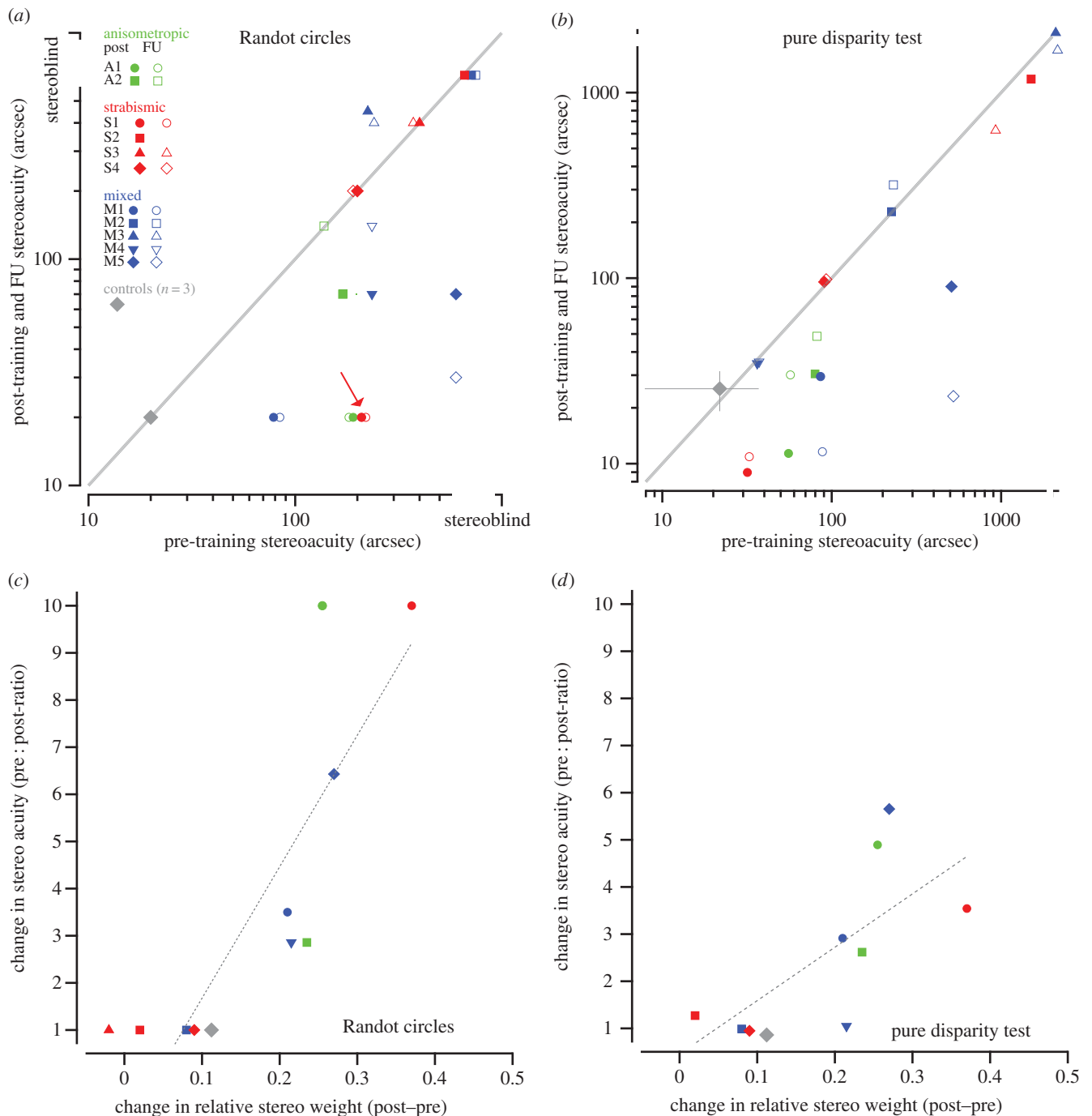
When they first began the task, our stereo-deficient subjects were quite poor at it as measured by the error in their hitting movements and a strong regression to the mean slant. One to six practice sessions were therefore used, depending on when accuracy and bias reached an asymptote. The number of practice sessions varied per subject, but data from the first three sessions with relatively stable performance were used to compute the pre-training weights. Owing to reduced variability in weight estimates with training, post-training (and follow-up) weights were based on data derived from two sessions after training.

The relative stereo weights of our 11 stereo-deficient subjects increased by, on average, a factor of  $3.07 \pm 0.85$  (figure 3*b* shows the mean data). In contrast, the three stereo-normal control observers showed only a factor of  $\approx 1.3$  change in relative stereo weights as training proceeded (figure 3*d*)—from  $0.35 \pm 0.05$  (average of the first three sessions) to  $0.46 \pm 0.01$  (average of the last two sessions). It is interesting to note that the feedback given in the experiment would have provided a signal not to weight the stereo cue any higher than 0.5.

A one-way repeated-measures omnibus ANOVA on log-transformed relative stereo weights (to offset violations of normality) was conducted to compare group performance at pre, post and follow-up. The results showed that subjects assigned significantly more weight to stereoscopic cues after training [ $F_{2,20} = 7.77$ ,  $p = 0.003$ ]. Bonferroni-adjusted pairwise comparisons showed increased stereo weights (decreased monocular cue weights) after training ( $p = 0.004$  one-tailed, mean change =  $0.15 \pm 0.04$ ), with the difference being still seen at follow-up ( $p = 0.028$  one-tailed, mean change =  $0.13 \pm 0.04$ ), indicating that the improvements were largely retained after two months of no intervention.

Figure 5 shows the post- versus pre-training (solid symbols) relative stereo weights for each of the 11 stereo-deficient observers. Points above the unity line indicate an increase in relative stereo weights following training. Eight of the 11 subjects showed a numerical increase in stereo weights following training, with these improvements being





**Figure 7.** (a) Randot circles test and (b) pure disparity test (PDT). Pre- versus post-training (solid symbols) and follow-up (open symbols) stereothresholds (Randot circles test, (a); PDT, (b)). Each coloured symbol shows the data of a single stereo-deficient observer. The grey diamond is the mean pre/post-data of the three control subjects who underwent training. Data under the diagonal unity line indicate decreases in stereothresholds (i.e. improved stereoacuity). The red arrow in (a) indicates the subject (S1), with the largest change in stereo weights. Note that small horizontal shifts have been applied to some of the data points to avoid symbols overlapping. (c) Randot circles test and (d) pure disparity test (PDT). The change in stereothresholds (pre : post-ratio) versus the change in relative stereo weights (post-pre) for the Randot circles test (c), and for the PDT (d), which has no monocular cues. The lines show the best-fitting linear regressions for the Randot circles and PDT, respectively. FU, follow up.

largely maintained at follow-up, two months later (except for two subjects who showed regression in gains). The inset in figure 5 shows the evolution of this increase in relative stereo weights for a strabismic subject (S1) who initially showed very low weighting of the stereo cues. A bootstrap analysis (10 000 iterations), performed to assess the intra-individual changes in relative stereo weights after training compared with their respective baselines indicates that the upweighting of the stereoscopic cues following training was statistically significant for six of the 11 stereo-deficient subjects at post-training (all  $p < 0.001$ ). In addition, two of

them (M2 and S4) showed the right trend from pre- to post-training to follow-up but failed to reach significance, in part owing to the high variance of the pre-test estimates.

### (c) Behavioural measures of suppression, stereoacuity and visual acuity following training

#### (i) Suppression

Suppression was substantially reduced following training. The Friedman statistical test was applied, because the data

were non-normally distributed. We found that the bug squashing training resulted in a significant change to levels of suppression (figure 6, Friedman test,  $\chi^2(2, n = 11) = 13.35$ ,  $p = 0.001$ ). Specifically, reduced suppression was found in all but one subject after training (Wilcoxon signed-ranks test,  $Z = -2.67$ ,  $p = 0.0025$  one-tailed,  $r = -0.57$ ; median change in interocular contrast ratios = 0.44). These effects were retained at follow-up compared with baseline ( $Z = -2.67$ ,  $p = 0.0025$  one-tailed,  $r = -0.57$ ; median change = 0.35), with no significant difference between the magnitudes of suppression measured at post-training versus at follow-up ( $Z = -1.46$ ,  $p = 0.12$  one-tailed,  $r = -0.31$ ; median change = 0.0).

### (ii) Stereoacuity

Importantly, bug squashing led to improvements not only on the trained task, but also in the untrained stereoacuity tasks. All six subjects who showed significant increases in relative stereo weights at post-test also demonstrated an improvement in stereoacuity as measured by the Randot circles test (figure 7*a*, solid symbols), and five of these six subjects also showed significant improvements in stereoacuity (all  $p$ -values < 0.01; 1000 bootstrap iterations were performed to assess intraindividual differences in pre- and post-stereothresholds) on the PDT (figure 7*b*). These improvements were largely maintained at follow-up two months after the cessation of training (figure 7*b*, open symbols), with the exception of two anisometropic amblyopes who either discontinued use of contact lenses post-training (A2), or used them only sparingly (A1). These two subjects demonstrated some regression of stereoscopic gains at follow-up. Finally, the two subjects who showed the right numerical trend in relative stereo weights as the experiment proceeded (M2 and S4) did not show stereoacuity transfer as measured either with the Randot or with the PDT.

*Randot circles test:* statistical tests were run on log stereo-thresholds. A value of 2.78 log arc seconds (or 600 arc seconds) was arbitrarily assigned if subjects failed the Randot circles test. Subjects unable to perform at that largest disparity were labelled as 'stereoblind'. There was a significant change in stereoacuity over time, Friedman test,  $\chi^2(2, n = 11) = 6.64$ ,  $p = 0.031$ . *Post hoc* comparisons showed that subjects exhibited a trend for improved stereoacuity after training (Wilcoxon,  $Z = -1.866$ ,  $p = 0.039$  one-tailed,  $r = -0.4$ ; median change = 0.46 log arc sec, equivalent to 65% median improvement), and at follow-up compared with pre-training baselines ( $Z = -2.106$ ,  $p = 0.02$  one-tailed,  $r = -0.45$ ; median change = 0.15 log arc sec, equivalent to 30% improvement).

*PDT:* a one-way repeated-measures ANOVA showed that stereoacuity as measured by the PDT improved significantly after training (main effect of time:  $F_{2,20} = 4.86$ ,  $p = 0.019$ ). *Post hoc* comparisons with Bonferroni correction showed that the stereoacuity improved post-training compared with pre-training baselines ( $p = 0.02$  one-tailed; mean change =  $0.28 \pm 0.09$  log arc sec, equivalent to  $35.7 \pm 10.9\%$  improvement), and this improvement was largely retained at follow-up ( $p = 0.046$  one-tailed; mean change at follow-up from pre-training =  $0.41 \pm 0.16$  log arc sec).

While these results show that learning a visuomotor task that focused on surface slant estimation transfers to stereo depth judgements, a stronger test for generalization would be the

ability to predict improvements on the transfer tests from the post-training changes in stereo weights. Figure 7*c,d* plot changes in stereo threshold (pre : post ratio) as measured by the Randot circles and PDT, respectively, against changes in stereo weights (post-pre relative stereo weights). The regression lines fit the data well, indicating that training-induced changes in stereo weights significantly predicted improvements on the Randot circles ( $\beta = 0.75$ ,  $t = 3.44$ ,  $p = 0.007$ ;  $R^2 = 0.57$ ), and PDT ( $\beta = 0.9$ ,  $t = 6.0$ ,  $p < 0.0001$ ;  $R^2 = 0.8$ ).

### (iii) Visual acuity

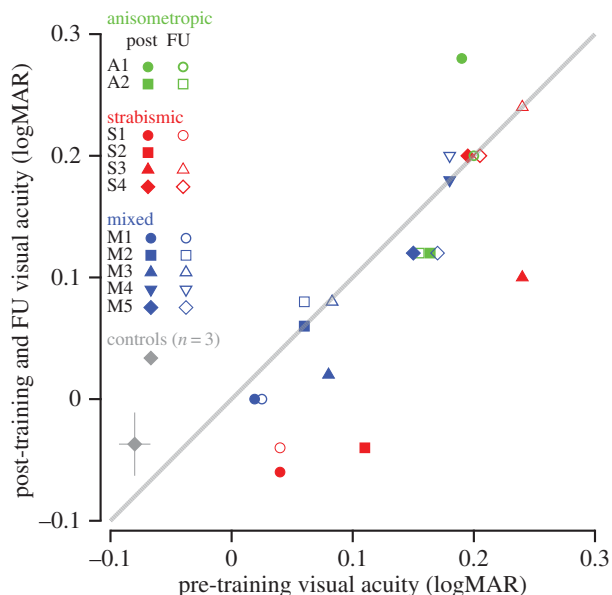
We examined whether direct stereo training might also offer training-induced benefits to visual acuity (figure 8). Six of the 11 subjects showed an improvement in visual acuity immediately post-training (greater than 0.04 logMAR), with two of these subjects regressing and one lost at follow-up. Of the three subjects that showed substantial (greater than or equal to 0.10 logMAR) improvements, two showed no post-training improvements in stereopsis and only one showed improvement in relative stereo weights, suggesting that changes in the non-dominant eye acuity are not tightly coupled to the training-related increases in relative stereo weights, in agreement with the results of Ding & Levi [8]. For the group as a whole, there was a trend for improved visual acuity in the non-dominant eye after training compared with the pre-training baseline; however, this effect was only marginally significant (repeated-measures ANOVA,  $F_{2,20} = 3.23$ ,  $p = 0.061$ ). Acuity of the dominant eye was unaltered post-training ( $F_{2,20} = 0.72$ ,  $p = 0.5$ ).

### (iv) Vergence

We measured vergence noise in eight of the participants (electronic supplemental material, figure S1). Prior to bug squashing training, the vergence noise in the experimental group was 4.9 times higher ( $9.72 \pm 3.5$  arc min,  $n = 8$ ) compared with the normally sighted subjects ( $1.97 \pm 0.2$  arc min,  $n = 4$ ). For the stereo-deficient group as a whole, there were no significant differences in vergence noise measured pre-training, post-training and at follow-up (Friedman test,  $\chi^2(2, n = 8) = 2.25$ ,  $p = 0.36$ ). Likewise, there were no changes in vergence bias with training (Friedman test,  $\chi^2(2, n = 8) = 3.25$ ,  $p = 0.24$ ). Four of these eight stereo-deficient subjects showed improvements in relative stereo weights following training, two of whom showed a post-training reduction of vergence noise, one showed an increase and one showed no change. There was a lack of consistency in the direction of change post-training. Thus, there is insufficient evidence to indicate that changes in the vergence noise can explain the training-related improvements in stereo weights in these subjects. We acknowledge that our vergence measure is a subjective measure, and thus may not be sufficiently sensitive to detect very small changes in oculomotor control ( $\approx 1.5$ – $2$  arc minutes in our normal control subjects—electronic supplementary material, figure S1).

### (v) Determinants of stereo weights improvements

We also examined other potential contributing factors to post-training changes in subjects' relative stereo weights. First, changes in interocular suppression failed to predict changes in stereo weights ( $\beta = -0.2$ ,  $t = -0.65$ ,  $p = 0.5$ ;  $R^2 = 0.04$ ). Second, we investigated whether any of the pre-training measures—logMAR acuity (non-dominant eye), interocular



**Figure 8.** Pre- versus post-training (solid symbols) and follow-up (open symbols) visual acuity (logMAR). Each coloured symbol shows the data of a single stereo-deficient observer. The grey diamond shows the mean pre/post-data of the three normal control observers who underwent training. Data under the diagonal unity line indicate increases in visual acuity. FU, follow up; logMAR, logarithm of the minimum angle of resolution.

difference in acuity, stereo weights, age and Randot and PDT stereosensitivities—could predict training-related changes in stereo weights. Multiple regression analysis revealed that only the initial PDT significantly predicted training-induced changes in stereo weights ( $\beta = -0.8$ ,  $t = -3.9$ ,  $p = 0.003$ ,  $R^2 = 0.63$ ). None of the other measures did ( $p > 0.05$ ). Subjects with the best stereo acuity on the PDTs at the outset showed the greatest increases in relative stereo weights, whereas those with no measurable stereopsis showed no change.

## 4. Discussion

While there have been a large number of studies of the effects of PL and videogame play in adults with amblyopia (for recent reviews, see [15–18]), only a few studies have trained stereopsis directly (for a review, [3]). Our results show that training in a natural visuomotor task in which some stimuli contained monocular texture as well as stereoscopic cues (cue-consistent trials), some contained only stereoscopic cues and some contained conflicting monocular and stereo cues, enabled adults with long-standing deficits in stereo vision to upweigh their reliance on stereoscopic cues (relative to monocular cues). Following training, participants showed improved accuracy for detecting slants in depth. Our simple Bayesian modelling indicates a very substantial reduction in stereoscopic cue noise over the course of training in stereo-deficient observers, along with a more modest reduction in monocular cues noise. The dramatic increase in stereoscopic cue reliability as training progressed is consistent with the increased reliance of the observers on stereoscopic cues (i.e. increased relative stereo weights). In addition, stereo-deficient observers also showed reduced suppression, significant improvement in stereoacuity and a weak trend for improved visual acuity.

Out of the 11 participants trained, six showed significantly greater reliance on stereoscopic cues as their training

on our VR bug squashing game progressed, with an additional two showing the right numerical trend. Of the six who showed a significant effect, all showed improved stereoacuity as measured by the Randot at pre- and post-training. Although only marginally significant, we note, as others have done before, that the Randot may lack the needed sensitivity to assess pre- to post-test improvements. For example, strabismic subject S1 (shown by the arrow in figures 5 and 7) achieved a stereoacuity of 20 arc sec, the lower limit of the Randot circles test—this is the subject shown in the inset of figure 5 whose relative stereo weights improved substantially over the course of training. We suspect that this floor (lowest tested value on the test) underestimates S1's stereoacuity, since she demonstrated a stereoacuity of  $\approx 9$  arc sec on the PDT. Five participants showed highly significant post-training improvements on the PDT (electronic supplementary material, table S1), and crucially, the best determinant of whether stereo weights improved was performance on the PDT at pre-test.

### (a) Use of different sensory cues

Adults with normal visual experience reduce sensory uncertainty by integrating information from different modalities (e.g. touch and vision [19,20]) and within the same modality (e.g. binocular disparity and texture information) for judging slant in depth [10,11,21]. However, observers deprived of normal binocular visual experience early in life have poor or absent stereopsis, and therefore rely on texture information for making judgements of surface slant. Thus, in our sample of observers with abnormal binocular vision, the pre-training relative stereo weight was, on average  $\approx 0.13 \pm 0.04$ . In contrast, normal control subjects had a relative stereo weight of  $0.39 \pm 0.05$ . Over the course of training, the relative stereo weights of the stereo-deficient subjects increased substantially. It is interesting to note that while the adult visual system is optimized for reducing sensory uncertainty by integrating texture and stereoscopic cues to slant, mature sensory integration is not evident until age 12 [22]. For infants and young children, the limitation is not owing to insensitivity to one cue but to an immaturity in sensory integration. In contrast, adults deprived of normal binocular vision are initially insensitive to the disparity information. With experience in bug squashing, the disparity information becomes more reliable. Importantly, our bug squashing training involves integrating not just multiple visual cues, but also the rich information from tactile and kinesthetic feedback [21].

### (b) Real-life impact

Following training, two participants reported better distance judgement during driving, and one was able to appreciate depth from autostereograms for the first time. Finally, one important practical outcome of our study is the finding that strabismic subjects who demonstrated post-training stereo-improvements were able to perform the clinical Randot test without the use of prisms, despite the presence of uncorrected ocular deviation. Compensating the deviation with prisms did not further improve stereo acuity in any of our strabismic subjects. We hypothesized that if this finding were true, then the stereo weights should similarly be unaffected by lack of compensation for deviation. To assess this, we measured stereo weights without prisms for six subjects who were initially trained on the bug squashing task with prisms. Our results showed that the difference between

relative stereo weights assessed with and without prisms was negligible (mean =  $0.01 \pm 0.02$ ).

Our strabismic subjects who recovered stereopsis had stereoacuities after training, as low as 9 arc sec (S1 on the PDT). How do we reconcile the improvement in stereopsis in the presence of an uncorrected strabismus? It has long been known that fusion is not a requirement for stereopsis. For example, observers with normal vision can correctly localize objects in depth even when they appear double [23]. Westheimer & Tanzman [24] demonstrated that observers with normal binocular vision can correctly identify the depth of diplopic stimuli with disparities up to about  $7^\circ$ . Blakemore [25] reported that the largest disparity that supported signed depth judgements was approximately  $12^\circ$ . More recently, Dengler & Kommerell [26] showed that normal subjects could distinguish binocularly disparate images from monocular double images with the same angular separation over large distances (up to  $21^\circ$ ), where one eye's target was presented in the fovea, and the other eye's target was presented in the periphery. This is similar to the situation in strabismus, and, as Dengler & Kommerell speculate, these long-range connections between the fovea of one eye and the periphery of the other may be the basis for anomalous correspondence in patients with strabismus. One potential issue with these studies is that subjects can assign a 'correct' disparity to an isolated monocular target in one eye, as though it were matched to an invisible target in the fovea [27,28]. Dengler & Kommerell, like the Foley *et al.* [29] study before it, did employ monocular targets as controls, where the subjects had to identify the monocular targets with a label 'on'. What is interesting is that many subjects did respond based on 'eye-of-origin' information, and so their data were discarded. Perhaps there were subtle differences that allowed some subjects to discriminate the targets that were supposed to be labelled 'crossed' from the targets that were supposed to be labelled 'on'. However, the main point is that there is evidence for at least coarse stereopsis for unfused stimuli in normal vision. Coarse stereopsis is considered to be important to extend the range of disparity sensitivity, as a guide to vergence eye movements, and as a 'back up' system for individuals with strabismus [30]. Indeed, there is recent evidence that coarse stereopsis develops much earlier than fine stereopsis [31], and that it may be relatively 'spared' in individuals with a history of amblyopia [32]. Whether the same mechanisms are capable of supporting stereoacuity better than one arc minute in strabismic patients or whether there are specialized mechanisms supporting anomalous correspondence is a matter for further research. For now, this study documents that it is possible for strabismic patients to recover stereoacuity as low as 9 arc sec.

### (c) Why does recovery of stereopsis require heroic methods?

Is not the rich real-world stimulation enough? We believe that there are several reasons why the natural environment is not sufficient. First, under normal binocular viewing, stimuli in the two eyes do not have equal perceived contrast, leading to suppression of the weak eye by the strong eye [18,33,34]. Moreover, for strabismic subjects, the two eyes' images fall on non-corresponding retinal areas, precluding normal fusion and triggering suppression. Thus, under normal

viewing, higher brain areas involved in sensory integration [35–38] receive weak and unreliable information from the weak eye, resulting in a down-weighting of disparity information. In contrast, in our bug squashing game, the stimuli to the two eyes were perceptually matched by reducing the contrast presented to the strong eye and were carefully aligned, so that the images can be fused. By combining this new aligned and balanced visual input with a visuomotor task (bug squashing), providing monocular depth cues as a scaffold, and giving trial-by-trial force feedback, subjects can learn to attend to input from the amblyopic eye [39] and can learn the correlations between the 'corrected' visual input and the depth of objects in the world. Finally, we should emphasize that our bug squashing training was conducted in a VR environment, enabling us to recreate 'the natural way in which perception and action are intimately entwined with the environment' [40, p. 2].

Our current results, taken together with previous studies (reviewed in [3]), suggest that to optimize stereo recovery it is critical to initially provide aligned and balanced input to the two eyes. This may require substantial fusion training [8] prior to direct stereo training in a rich environment that combines a natural visuomotor task.

## 5. Summary and conclusion

We trained adults who were stereoblind or stereo-deficient on a natural visuomotor 'bug squashing' task in which some stimuli contained monocular texture cues as well as stereoscopic cues that were consistent with each other, some contained only stereoscopic cues and some contained conflicting monocular and stereoscopic cues. Following training, eight out of 11 observers upweighted their reliance on stereoscopic cues (relative to monocular cues), and six of them showed reduced suppression and significant improvements in stereoacuity.

Our results have important implications for the recovery of visual function late in life, well outside the 'critical period', which until recently was thought to offer the only real scope for plasticity. More broadly, our approach demonstrates the potential power offered by VR for perceptual training of all kinds.

**Ethics.** The study was approved by the Institutional Review Board at the University of Rochester, and informed consent was obtained from all subjects.

**Data accessibility.** The datasets supporting this article have been uploaded as part of the electronic supplementary material.

**Authors' contributions.** The initial conceptual framework for the VR training was developed by D.C.K. D.C.K., D.M.L. and D.B. designed the study and obtained a research grant from the National Eye Institute. Adapting the VR training paradigm for clinical application: I.V. for design aspects and feedback, and A.Y. & D.C.K. for the programming. Piloting and fine-tuning of the training and vision assessments, primarily I.V. with contributions from A.Y., S.J.H. and D.C.K.; running the study: primarily, I.V. with contributions from S.J.H. and the supervision of D.B.; data analysis: primarily I.V. with the help of D.C.K. and D.B.; modelling of stereo weight data: O.-S.K.; J.D. developed the PDT and assisted with its data analysis. Writing: primarily I.V. and D.M.L. with the feedback of D.C.K. and D.B., and other authors contributing as needed. Final figures: D.M.L.; D.B. and D.M.L. are co-senior authors.

**Competing interests.** We have no competing interests.

**Funding.** This research was supported by grants RO1EY020976 from the National Eye Institute to D.M.L., D.C.K. and D.B.

**Acknowledgements.** Our dear friend and colleague David Knill died tragically last year. We, and indeed the entire vision community, will miss

him sorely, and we dedicate this paper to his memory. We thank our research assistants and study coordinators Olga Pikul, Leslie Chylinsky, Aleksandra Fazlipour and Gary Volkell. We also thank

Dr Hu for assistance with programming. We thank Drs Gearinger and DePaolis for referring patients to our study. Finally, this work could not have happened without the cooperation of our subjects.

## References

- Tam WJ, Stelmach LB. 1998 Display duration and stereoscopic depth discrimination. *Can. J. Exp. Psychol.* **52**, 56–61. (doi:10.1037/h0087280)
- Zaroff CM, Knutelska M, Frumkes TE. 2003 Variation in stereoacuity: normative description, fixation disparity, and the roles of aging and gender. *Invest. Ophthalmol. Vis. Sci.* **44**, 891–900. (doi:10.1167/iovs.02-0361)
- Levi DM, Knill DC, Bavelier D. 2015 Stereopsis and amblyopia: a mini-review. *Vision Res.* **114** 17–30. (doi:10.1016/j.visres.2015.01.002)
- Barry SO. 2009 *Fixing my gaze: a scientist's journey into seeing in three dimensions*. New York, NY: Basic Books.
- Bridgeman B. 2014 Restoring adult stereopsis: a vision researcher's personal experience. *Optom. Vis. Sci.* **91**, e135–e139. (doi:10.1097/OPX.0000000000000272)
- Nakatsuka C, Zhang B, Watanabe I, Zheng J, Bi H, Ganz L, Smith EL, Harwerth RS, Chino YM. 2007 Effects of perceptual learning on local stereopsis and neuronal responses of V1 and V2 in prism-reared monkeys. *J. Neurophysiol.* **97**, 2612–2626. (doi:10.1152/jn.01001.2006)
- Astle AT, McGraw PV, Webb BS. 2011 Recovery of stereo acuity in adults with amblyopia. *BMJ Case Rep.* **2011**, 1–4. (doi:10.1136/bcr.07.2010.3143)
- Ding J, Levi DM. 2011 Recovery of stereopsis through perceptual learning in human adults with abnormal binocular vision. *Proc. Natl Acad. Sci. USA* **108**, E733–E741. (doi:10.1073/pnas.1105183108)
- Xi J, Jia WL, Feng LX, Lu ZL, Huang CB. 2014 Perceptual learning improves stereoacuity in amblyopia. *Invest. Ophthalmol. Vis. Sci.* **55**, 2384–2391. (doi:10.1167/iovs.13-12627)
- Knill DC. 2005 Reaching for visual cues to depth: the brain combines depth cues differently for motor control and perception. *J. Vis.* **5**, 103–115. (doi:10.1167/5.2.2)
- Knill DC, Saunders JA. 2003 Do humans optimally integrate stereo and texture information for judgments of slant? *Vis. Res.* **43**, 2539–2558. (doi:10.1016/S0042-6989(03)00458-9)
- Simons K. 1981 A comparison of the Frisby, random-dot E, TNO, and Randot circles stereotests in screening and office use. *Arch. Ophthalmol.* **99**, 446–452. (doi:10.1001/archoph.1981.03930010448011)
- McKee SP, Levi DM. 1987 Dichoptic hyperacuity: the precision of nonius alignment. *J. Opt. Soc. Am. A* **4**, 1104–1108. (doi:10.1364/JOSAA.4.001104)
- Acerbi L, Wolpert DM, Vijayakumar S. 2012 Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Comput. Biol.* **8**, e1002771. (doi:10.1371/journal.pcbi.1002771)
- Levi DM, Li RW. 2009 Perceptual learning as a potential treatment for amblyopia: a mini-review. *Vision Res.* **49**, 2535–2549. (doi:10.1016/j.visres.2009.02.010)
- Levi DM. 2012 Prentice award lecture 2011: removing the brakes on plasticity in the amblyopic brain. *Optom. Vis. Sci.* **89**, 827–838. (doi:10.1097/OPX.0b013e318257a187)
- Hess RF, Thompson B. 2015 Amblyopia and the binocular approach to its therapy. *Vision Res.* **114** 4–16. (doi:10.1016/j.visres.2015.02.009)
- Tsirlin I, Colpa L, Goltz HC, Wong AM. 2015 Behavioral training as new treatment for adult amblyopia: a meta-analysis and systematic review. *Invest. Ophthalmol. Vis. Sci.* **56**, 4061–4075. (doi:10.1167/iovs.15-16583)
- Ernst MO, Banks MS. 2002 humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433. (doi:10.1038/415429a)
- Ernst MO, Banks MS, Bulthoff HH. 2000 Touch can change visual slant perception. *Nat. Neurosci.* **3**, 69–73. (doi:10.1038/71140)
- Knill DC, Kersten D. 2004 Visuomotor sensitivity to visual information about surface orientation. *J. Neurophysiol.* **91**, 1350–1366. (doi:10.1152/jn.00184.2003)
- Nardini M, Bedford R, Mareschal D. 2010 Fusion of visual cues is not mandatory in children. *Proc. Natl Acad. Sci. USA* **107**, 17 041–17 046. (doi:10.1073/pnas.1001699107)
- Ogle KN. 1952 Disparity limits of stereopsis. *AMA Arch. Ophthalmol.* **48**, 50–60. (doi:10.1001/archoph.1952.00920010053008)
- Westheimer G, Tanzman IJ. 1956 Qualitative depth localization with diplopic images. *J. Opt. Soc. Am.* **46**, 116–117. (doi:10.1364/JOSA.46.000116)
- Blakemore C. 1970 The range and scope of binocular depth discrimination in man. *J. Physiol.* **211** 599–622. (doi:10.1113/jphysiol.1970.sp009296)
- Dengler B, Kommerell G. 1993 Stereoscopic cooperation between the fovea of one eye and the periphery of the other eye at large disparities. Implications for anomalous retinal correspondence in strabismus. *Graefes Arch. Clin. Exp. Ophthalmol.* **231**, 199–206. (doi:10.1007/BF00918841)
- Kaye M. 1978 Stereopsis without binocular correlation. *Vision Res.* **18**, 1013–1022. (doi:10.1016/0042-6989(78)90029-9)
- Wilcox LM, Harris JM, McKee SP. 2007 The role of binocular stereopsis in monoptic depth perception. *Vision Res.* **47**, 2367–2377. (doi:10.1016/j.visres.2007.04.022)
- Foley JM, Applebaum TH, Richards WA. 1975 Stereopsis with large disparities: discrimination and depth magnitude. *Vision Res.* **15**, 417–421. (doi:10.1016/0042-6989(75)90091-7)
- Wilcox LM, Allison RS. 2009 Coarse–fine dichotomies in human stereopsis. *Vision Res.* **49**, 2653–2665. (doi:10.1016/j.visres.2009.06.004)
- Giaschi D, Narasimhan S, Solski A, Harrison E, Wilcox LM. 2013 On the typical development of stereopsis: fine and coarse processing. *Vision Res.* **89** 65–71. (doi:10.1016/j.visres.2013.07.011)
- Giaschi D, Lo R, Narasimhan S, Lyons C, Wilcox LM. 2013 Spacing of coarse stereopsis in stereo-deficient children with a history of amblyopia. *J. Vis.* **13**, pii. 17. (doi:10.1167/13.10.17)
- Ding J, Klein SA, Levi DM. 2013 Binocular combination in abnormal binocular vision. *J. Vis.* **13**, 14. (doi:10.1167/13.2.14)
- Vedamurthy I, Nahum M, Bavelier D, Levi DM. 2015 Mechanisms of recovery of visual function in adult amblyopia through a tailored action video game. *Sci. Rep.* **5**, 8482. (doi:10.1038/srep08482)
- Welchman AE, Deubelius A, Conrad V, Bulthoff HH, Kourtzi Z. 2005 3D shape perception from combined depth cues in human visual cortex. *Nat. Neurosci.* **8**, 820–827. (doi:10.1038/nn1461)
- Morgan ML, Deangelis GC, Angelaki DE. 2008 Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* **59**, 662–673. (doi:10.1016/j.neuron.2008.06.024)
- Ban H, Preston TJ, Meeson A, Welchman AE. 2012 The integration of motion and disparity cues to depth in dorsal visual cortex. *Nat. Neurosci.* **15**, 636–643. (doi:10.1038/nn.3046)
- Murphy AP, Ban H, Welchman AE. 2013 Integration of texture and disparity cues to surface slant in dorsal visual cortex. *J. Neurophysiol.* **110**, 190–203. (doi:10.1152/jn.01055.2012)
- Li RW, Ngo CV, Levi DM. 2015 Relieving the attentional blink in the amblyopic brain with video games. *Sci. Rep.* **5**, 8483. (doi:10.1038/srep08483)
- Scarfe P, Glennerster A. 2015 Using high-fidelity virtual reality to study perception in freely moving observers. *J. Vis.* **15**, 3. (doi:10.1167/15.9.3)