Supplementary materials for:

Visual acuity, crowding and stereo-vision are linked in children with and without amblyopia

John A. Greenwood[1,2,3,4], Vijay K. Tailor[2,5], John J. Sloper[2,5], Anita J. Simmers[6], Peter J. Bex[7], and Steven C. Dakin[1,2]

[1] *UCL Institute of Ophthalmology, University College London, London, UK*
[2] *NIHR Biomedical Research Centre for Ophthalmology at Moorfields Eye Hospital NHS Foundation Trust, London, UK*
[3] *Laboratoire Psychologie de la Perception, Université Paris Descartes, Sorbonne Paris Cité, Paris, France*
[4] *Centre National de la Recherche Scientifique, UMR 8158, Paris, France*
[5] *Strabismus and Paediatric Service, Moorfields Eye Hospital, London, UK*
[6] *Vision Sciences, Department of Life Sciences, Glasgow Caledonian University, Glasgow, UK*
[7] *Schepens Eye Research Institute, Harvard Medical School, Boston, MA, USA*

Supplementary movie files

We include four supplementary movie files to demonstrate the operation of each task in the VacMan battery. For demonstration, screen dimensions have been reduced to 640×480 pixels, though this was considerably larger in the actual task (see main text). A description of each movie follows:

Supplementary Movie 1. Two example trials of the acuity task. The first response is incorrect, causing an increase in stimulus size on the second trial. The second response is correct and the reward animation follows (though in the actual task, three correct responses were required for the animation).

Supplementary Movie 2. Two example trials of the contrast-detection task. The first response is correct and VacMan becomes dimmer on the second trial as a result.

Supplementary Movie 3. Two example trials of the crowding task. The first response is incorrect, and the centre-to-centre spacing between VacMan and the achromatic "ghost" flankers increases on the subsequent trial.

Supplementary Movie 4. Two example trials of the stereo-acuity task, with monocular "shadows" present. The presence of binocular disparity is demonstrated here by showing the signal to each eye on subsequent frames of the movie (causing the "ghost" with disparity to move back-and-forth, which was not present in the actual task). The first trial elicits a correct response, causing the binocular disparity to decrease on the subsequent trial.

Clinical details of the children

All children underwent a full orthoptic examination prior to participating in the experiments. The four tables on the following pages display the details of each child and their results in the orthoptic exam. The selection criteria for each group are described in-text, but note that there was no requirement for the presence of amblyopia in the three "clinical" groups, as we wished to assess a range of visual abilities. Though the majority do meet this criterion (either acuity worse than 0.1 logMAR or an interocular acuity difference greater than 0.1 logMAR), one of the anisometropic children (MF) and four of the strabismic children (SS, AI, JM, and HA) do not. Nonetheless, removing these children does not qualitatively alter the analyses described in-text. We also sought a mix of stereo-vision abilities for each group, and thus the observed frequencies of stereo-blindness for these groups should not be taken as indicative of the overall population (e.g. Richards, 1970, 1971).

Initials	Age (months)	Sex	Refractive error	logMAR acuity	TNO stereo.
MW	67	F	R: Plano L: Plano	R: 0.00 L: 0.00	60″
KK	70	F	R: +2.00/-1.25×180° L: +2.00/-1.25×180°	R: 0.075 L: 0.075	240″
OC	64	F	R: Plano L: Plano	R: 0.06 L: 0.02	60″
IB	85	M	R: Plano L: Plano	R: 0.00 L: 0.00	120″
BM	94	F	R: Plano L: Plano	R: 0.00 L: 0.00	60″
BR	83	F	R: Plano L: Plano	R: 0.05 L: 0.00	60″
AH	86	M	R: Plano L: Plano	R: 0.025 L: 0.025	60″
FC	75	M	R: Plano L: Plano	R: 0.04 L: 0.04	60″
KA	63	M	R: Plano L: Plano	R: 0.00 L: 0.00	120″
BD	83	F	R: Plano L: Plano	R: 0.00 L: 0.00	60″
SA	90	M	R: Plano L: Plano	R: -0.10 L: 0.00	60″
SL	75	M	R: Plano L: Plano	R: 0.025 L: 0.075	120″
JB	70	M	R: Plano L: Plano	R: 0.00 L: 0.00	60″
SA	92	F	R: Plano L: Plano	R: 0.06 L: 0.00	60″
JT	100	F	R: +2.75 DS L: +2.75/-0.50×165°	R: 0.00 L: -0.04	60″
IK	86	F	R: Plano L: Plano	R: 0.00 L: 0.00	60″
GJ	60	M	R: Plano L: Plano	R: -0.10 L: -0.10	60″
EG	60	F	R: Plano L: Plano	R: 0.00 L: 0.04	60″
GF	81	F	R: Plano L: Plano	R: 0.00 L: 0.00	60″

Supplementary Table 1. Clinical details of the 19 control children. Age is reported in months. Optical correction includes cylindrical and spherical values with appropriate axes for each eye, where R = right eye and L = left eye. logMAR acuity is similarly reported for each eye, and results of the TNO stereo-acuity test are reported in seconds of arc.

Initials	Age (months)	Sex	Refractive error	logMAR acuity	TNO stereo.
GF	102	M	R: +7.00/-1.00×180° L: +6.00/-1.00×180°	R: 0.10 L: 0.10	360″
AK	83	F	R: +1.00/-0.5×5° L: +3.50/-0.5×180°	R: 0.00 L: 0.125	60″
SA	85	F	R: -5.75/-0.5×20° L: +1.75/-1.25×20°	R: 0.25 L: 0.05	Nil
RH	101	M	R: +4.50 DS L: +5.50/-2.25×175	R: 0.025 L: 0.125	120″
AV	98	F	R: +1.25 DS L: +7.00/-1.00×180°	R: -0.22 L: 0.10	240″
AA	80	M	R: -0.25/-1×25° L: 0/-2.00×165°	R: 0.00 L: 0.10	60″
BH	72	M	R: Plano L: +4.50 DS	R: -0.10 L: 0.10	120″
CT	95	M	R: +6.50/-0.50×60° L:+1.75 DS	R: 0.10 L: -0.10	240″
JK	80	F	R:+2.50/-0.25×180° L:+6.00/-1.75×170°	R: 0.00 L: 0.10	120″
JB	86	M	R:+1.00 DS L:+3.00 DS	R: 0.00 L: 0.28	120″
NR	104	F	R:+5.50/-1.25×180° L:+1.00/-0.50×5°	R: 0.60 L: 0.00	Nil
PS	90	F	R:+5.75/-1.75×10° L:+2.75/-0.50×170°	R: 0.14 L: 0.00	120″
MF	83	F	R: -0.50/-0.50×180° L:+1.25/-0.25×180°	R: 0.00 L: 0.00	60″
CB	62	M	R:+3.25/-1.5×10° L:+2.00/-0.75×180°	R: 0.80 L: 0.10	Nil
LC	108	M	R:+5.50 DS L:+3.75 DS	R: 0.12 L: -0.10	60″
SS	56	F	R:+3.25/-3.50×95° L:+1.75/-1.00×95°	R: 0.20 L: 0.00	160″

Supplementary Table 2. Clinical details of the 16 anisometropic children. Values reported are in the same format as Supplementary Table 1.

Initials	Age (months)	Sex	Ocular alignment (with Rx)	Refractive error	logMAR acuity	TNO stereo.
MO	80	F	n: L Alt. SOT 16Δ d: L Alt. SOT 6Δ	R: +5.50/-3.00×180° L: +5.75/-2.50×165°	R: 0.275 L: 0.45	Nil
SS	92	F	n: straight d: R XOT 18Δ	R: Plano L: Plano	R: 0.00 L: 0.00	60″
NJ	94	M	n: L XOT 10Δ d: L XOT 8Δ L/R 10Δ	R: +3.75/-0.25×10° L: +4.50/-0.25×70°	R: 0.10 L: 0.15	Nil
EG	60	F	n: straight d: R XOT 20Δ	R: +1.00/-0.50×90° L: +1.50 DS	R: 0.15 L: 0.00	240″
GD	92	F	n: L SOT 45Δ d: L SOT 40Δ	R: +5.75/-1.25×175° L: +6.50/-1.25×180°	R: 0.15 L: 0.75	Nil
DS	94	M	n: R SOT 8Δ R/L 6Δ d: R SOT 6Δ R/L 11Δ	R: Plano L: Plano	R: 0.20 L: 0.00	Nil
HD	83	M	n: R SOT 10Δ d: R SOT 6Δ	R: +8.50/-0.50×155° L: +8.25/-2.00×45°	R: 0.36 L: 0.10	Nil
JB	64	M	L SOT 20Δ	R: +4.00/+0.50×180° L: +4.75/-0.75×175°	R: 0.06 L: 0.46	Nil
NO	96	M	L SOT <10Δ	R: +1.50 DS L: +2.25 DS	R: -0.14 L: 0.06	60″
OC	105	M	R SOT 12Δ	R: +7.00/-1.00×20° L: +7.50/-1.50×170°	R: 0.54 L: 0.10	Nil
RS	75	M	L SOT 40Δ	R: +8.25/-1.00×30° L: +8.25/-1.50×160°	R: 0.10 L: 0.76	Nil
LH	60	F	L SOT <10Δ	R: +6.25/-0.75×180° L: +6.50 DS	R: 0.00 L: 0.18	Nil
AI	53	F	L SOT <10Δ	R: +5.50/-0.50×180° L: +5.75/-0.25×30°	R: 0.00 L: 0.06	Nil
TP	95	F	R SOT 10Δ	R: +6.50/-2.50×140° L: +6.50/-2.75×170°	R: 0.40 L: 0.10	Nil
CA	97	F	n: L SOT 45Δ d: L SOT 25Δ	R: +2.50 DS L: +2.25 DS	R: 0.10 L: 0.15	Nil
JM	58	M	Alt. R SOT 15Δ	R: +1.50 DS L: +1.25 DS	R: 0.06 L: 0.02	Nil
HA	61	F	n: Alt. R SOT 20Δ d: Alt. R SOT 10Δ	R: +4.50 DS L: +5.00 DS	R: 0.02 L: 0.04	60″
HD	61	F	R SOT 20Δ	R: +6.00 DS L: +6.00 DS	R: 0.26 L: 0.00	Nil

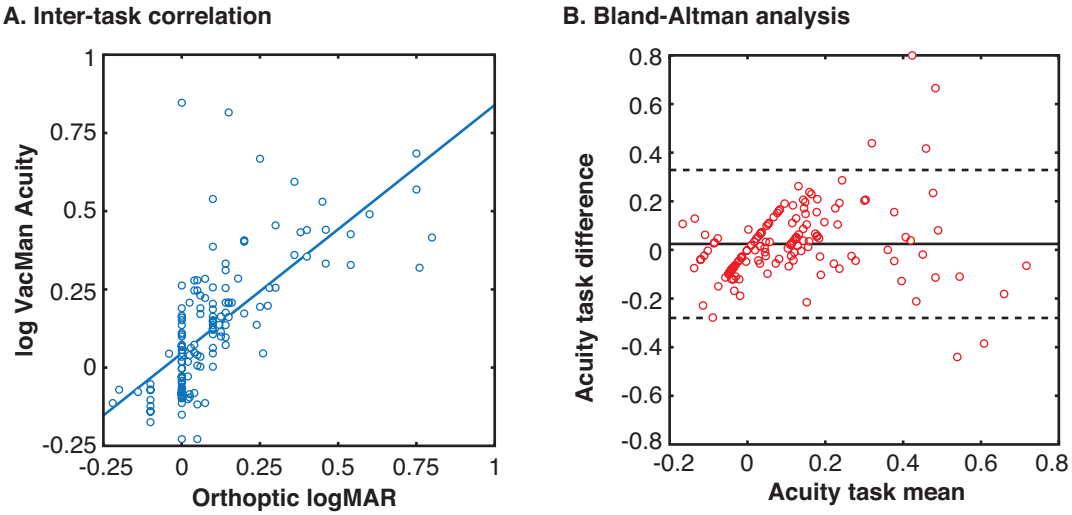
Supplementary Table 3. Clinical details of the 18 strabismic children. The "ocular alignment" column reports the outcome of both near (n) and distance (d) prism cover tests. Here, SOT = esotropia, XOT = exotropia, Alt. = alternating, L/R = left over right, and R/L = right over left. The degree of deviation is shown in prism dioptres and the amblyopic eye is denoted. All other values are in the same format as the preceding tables.

Initials	Age (months)	Sex	Ocular alignment (with Rx)	Refractive error	logMAR acuity	TNO stereo.
JP	86	M	R SOT 10Δ	R: -9.75/-1.50×75° L -3.25/-1.75×125°	R: 0.75 L: 0.05	Nil
SH	95	M	L SOT <10Δ	R: +3.25 DS L: +6.75 DS	R: -0.10 L: 0.05	120″
AW	65	M	n: R SOT 25Δ d: R SOT 20Δ	R: +7.25/-1.50×20° L: +6.25/-1.50×155°	R: 0.25 L: 0.05	Nil
RM	70	M	L SOT 25Δ L/R 12Δ	R: +2.50/-0.50×180° L: +4.00/-1.75×65°	R: 0.05 L: 0.125	Nil
GE	75	F	n: R SOT 16Δ d: R SOT 12Δ	R: +4.50/-0.5×160° L: +0.75 DS	R: 0.20 L: 0.00	Nil
BW	82	M	L SOT 30Δ	R: +3.00/-1.00×100° L: +6.00/-1.50×80°	R: -0.20 L: 0.14	Nil
AJ	96	M	L SOT <10Δ	R: 0.00/-0.50×5° L: +4.00/-1.25×175°	R: 0.00 L: 0.16	240″
MM	59	F	R SOT 30Δ	R: +6.25/-1.25×115° L: +5.25/-0.75×50°	R: 0.30 L: 0.10	Nil
AW	66	M	R SOT 35Δ	R: +7.00/-0.75×180° L: +6.00/-1.00×180°	R: 0.38 L: 0.10	Nil
EO	62	F	L SOT 10Δ	R: +4.5 DS L: +6.5 DS	R: 0.10 L: 0.36	Nil
SG	104	F	R SOT <10Δ	R: +5.50/-4.00×10° L: +3.00/-1.50×160°	R: 0.14 L: -0.10	60″
DB	84	F	L SOT <10Δ	R: +2.75 DS L: +4.25 DS	R: 0.00 L: 0.30	360″
EP	80	F	L SOT <10Δ	R: +1.5/-0.5×180° L: +6.25/-2.25×10°	R: 0.00 L: 0.40	Nil
TH	107	M	L SOT 12Δ L/R 9Δ	R: +1.75/-0.5×20° L: +3.00/-1.5×170°	R: 0.00 L: 0.54	Nil
HB	70	F	L SOT <10Δ	R: +4.25/-0.50×150° L: +5.25/-0.75×170°	R: 0.14 L: 0.46	Nil
NO	77	F	R SOT 18Δ	R: +4.5/-0.25×180° L: +3.5/-0.50×180°	R: 0.14 L: 0.03	Nil
NT	73	F	L SOT <10Δ	R: +4.25 DS L: +6.25 DS	R: 0.04 L: 0.14	450″
DB	93	M	n: R SOT 25Δ d: R SOT 14Δ	R: +4.00 DS L: +2.75/+0.75×180°	R: 0.24 L: 0.00	Nil
JD	77	F	SOT 30Δ	R: +3.50/-0.75×5° L: +5.00/+3.50×175°	R: 0.14 L: 0.14	Nil

Supplementary Table 4. Clinical details of the 19 mixed anisometropic and strabismic children. All values are reported in the same format as Supplementary Table 3.

Inter-task correlation and agreement

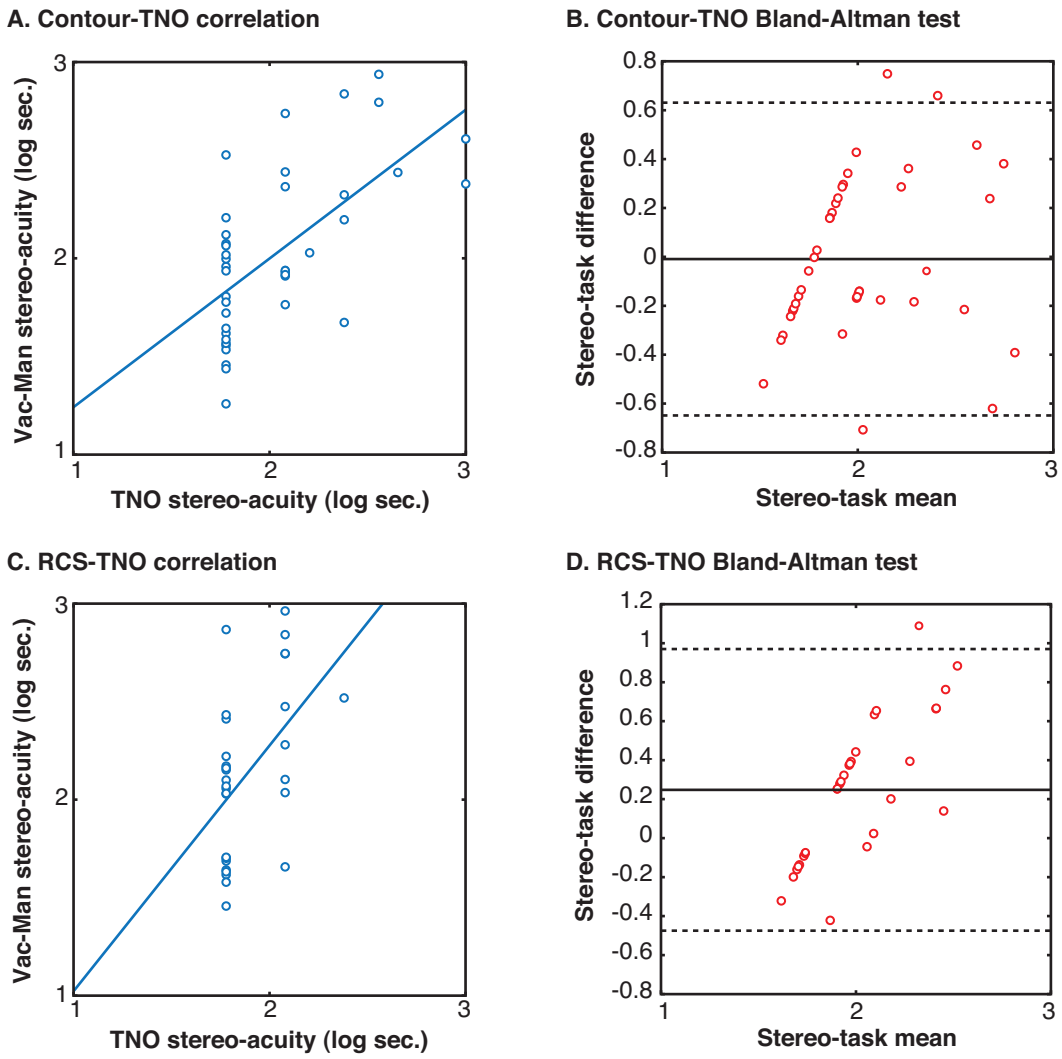
Because all children underwent a full orthoptic examination prior to their inclusion in the study, we can compare values obtained using our *Vac-Man* procedures and those obtained using more established methods. In the clinic, logMAR acuity was measured with the Thompson V2000. When these values are compared with our acuity thresholds (in equivalent units of log minutes of arc), a strong correlation is observed: $r(142) = 0.68$, $p<.001$ (Supplementary Figure 1A). A Bland-Altman analysis (Bland & Altman, 1986) was conducted to examine the level of agreement between these values. This involves a comparison between the mean of acuity values on each task as a function of the difference between these values, as plotted in Supplementary Figure 1b. This shows an acceptable level of agreement between the two tasks: 95% of our acuity values were between ±0.31 log min. (dashed lines), around a mean difference of 0.02 log minutes.



Supplementary Figure 1. A. Correlation between our VacMan acuity thresholds and logMAR values obtained during orthoptic examination. Our values (in log minutes of arc) correlate strongly with orthoptic values. **B.** Bland-Altman test between VacMan acuity thresholds and the orthoptic logMAR values. The x-axis plots the average of the orthoptic logMAR values and log VacMan Acuity values, while the y-axis plots the difference between the two values. Dashed lines indicate the 95% confidence limits on the differences between values, which indicate that the differences largely fall within a range of ±0.31 log minutes around a mean difference (solid line) of 0.02 log minutes.

Stereo-acuity was also assessed during the orthoptic examinations, using the TNO circles test. For comparison with our Vac-Man measures, we converted these values into log seconds of arc. If we examine only those with measurable stereo-thresholds on the contour-based Vac-Man stereo task, there is a strong correlation between these two

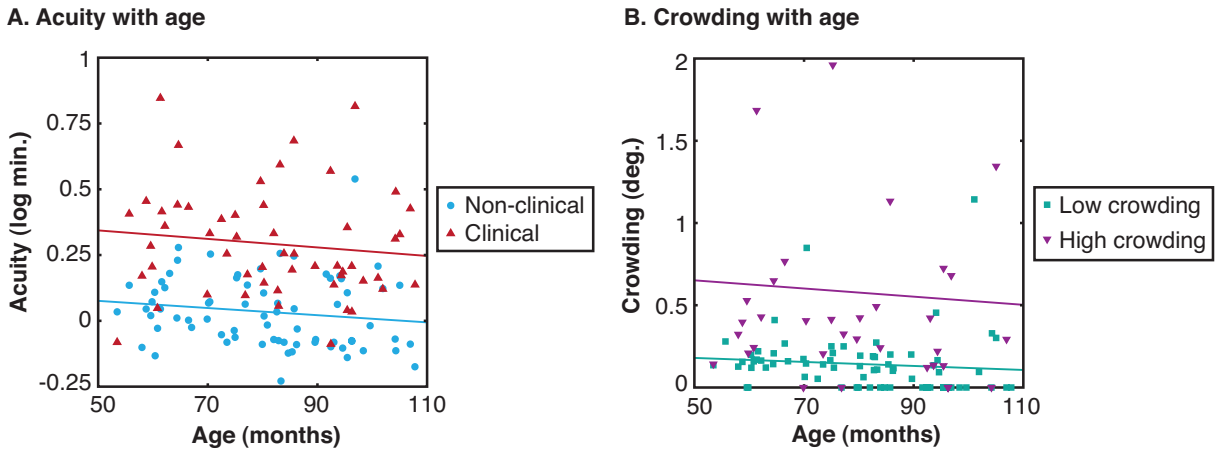
measures ($r(39)=0.63$ $p<0.001$), as plotted in Supplementary Figure 2A. When submitted to a Bland-Altman analysis, as in Supplementary Figure 2B, these values show modest agreement, with 95% of values falling within ±0.64 log seconds around a mean of -0.01 log seconds. Similarly, for the random-check stereogram (RCS) version of the task, there was a strong correlation with TNO performance ($r(28)=0.49$ $p=0.01$), as plotted in Supplementary Figure 2C, despite the lower number of children with measurable thresholds in this task. The Bland-Altman analysis for these tasks, plotted in Supplementary Figure 2D, reveals some degree of bias – Vac-Man RCS thresholds were consistently higher than those obtained in the TNO, with a mean difference of 0.25 log seconds. The spread of error is nonetheless similar, with 95% of values contained within ±0.72 log seconds. The discrepancy between these stereo-measures is likely due to our Vac-Man task being conducted at a distance of 3m, compared to the arm's length distance of the TNO test. Although this greater distance has the advantage of minimising potential motion parallax cues (Ono, Rivest, & Ono, 1986), it also increases the likelihood that individual checks in the display will fall below acuity limits and fail to be resolved. The same is not true for the contour-based task where lower spatial frequency information is also available. Future experiments using the Vac-Man stereo-task may nonetheless benefit from closer viewing distances.



Supplementary Figure 2. A. Correlation between our Vac-Man stereo-acuity thresholds, obtained with monocular "shadows" present, and values obtained from the TNO circles test during orthoptic examination. Both values (in log seconds of arc) correlate strongly. **B.** Bland-Altman test between our contour-based stereo-acuity thresholds and the TNO values. The x-axis plots the average of the orthoptic two values, while the y-axis plots the difference. Dashed lines indicate the 95% confidence limits on the differences between values. **C**. Correlation between our Vac-Man stereo-acuity thresholds, obtained in the random-check stereogram (RCS) version of the task, and TNO values, plotted as in panel A. **D**. Bland-Altman test between our random-check stereogram Vac-Man test and the TNO circles test, plotted as in panel B.

The effect of age

We include in our sample of children an age range from 54-107 months, with a mean of 80.7 months. To examine whether this age range contains variation in abilities, we examined the effect of age on acuity and crowding. Prior studies have shown that acuity is typically adult-like by approximately 6 years of age (Atkinson, Anker, Evans, & McIntyre, 1987; Ellemberg, Lewis, Hong Liu, & Maurer, 1999; Pan et al., 2009; Jeon, Hamid, Maurer, & Lewis, 2010), approximately midway through our age-range. The improvement over time for amblyopic children will obviously be confounded by the age at which treatment regimes begin, and so we separate amblyopic eyes from the fellow eyes and those of controls. Supplementary Figure 3A plots the relationship between acuity and age (in months) for these two groups of values. The "non-amblyopic" group includes averaged values for each control subject, and the fellow eye of all our amblyopic children. Computed in this way, there is no relationship between acuity and age over this range, $r(70)=-.15$, $p=.22$. Similarly, there is no improvement in acuity for the amblyopic eyes of children in the three clinical groups, $r(51)=-.12$, $p=.38$.



Supplementary Figure 3. Age effects on acuity and crowding. **A.** The effect of age (in months) on acuity values (in log minutes of arc). Here, "non-amblyopic" values (blue circles) represent the average of both eyes (for control subjects) or the fellow eye of children in the three clinical groups. The "amblyopic" values (red triangles) represent the amblyopic eye of children in all three clinical groups. There is no relationship with age for either group. **B.** The effect of age on crowding (in degrees). Here the "low crowding" values (green squares) represent the average of both eyes (for control subjects) or the fellow eye of children in the three clinical groups, while the "high crowding" values are the amblyopic eyes of children in the strabismic and mixed groups. There is no relationship for either dataset.

In contrast with the early development of acuity, elevated foveal crowding has been observed in "normally" developing children as late as 11 years (Atkinson & Braddick,

1983; Atkinson et al., 1987; Jeon et al., 2010). However, it is possible that age may improve crowding in the amblyopic group due to the progression of treatment in this time. To examine this, data was separated into those with "high crowding", i.e. the amblyopic eye of children in the strabismic and mixed groups, with comparison to those with "low crowding" – an average of foveal values in control children, and the fellow eye of all amblyopes. When computed this way, as plotted in Supplementary Figure 3B, there is no relationship between age and crowding (in degrees) for either of the "low crowding" ($r(70)=-0.10$, $p=.41$) or "high crowding" ($r(35)=-0.06$, $p=.72$) datasets.

Correlations between stereo-acuity thresholds and the spatial tasks

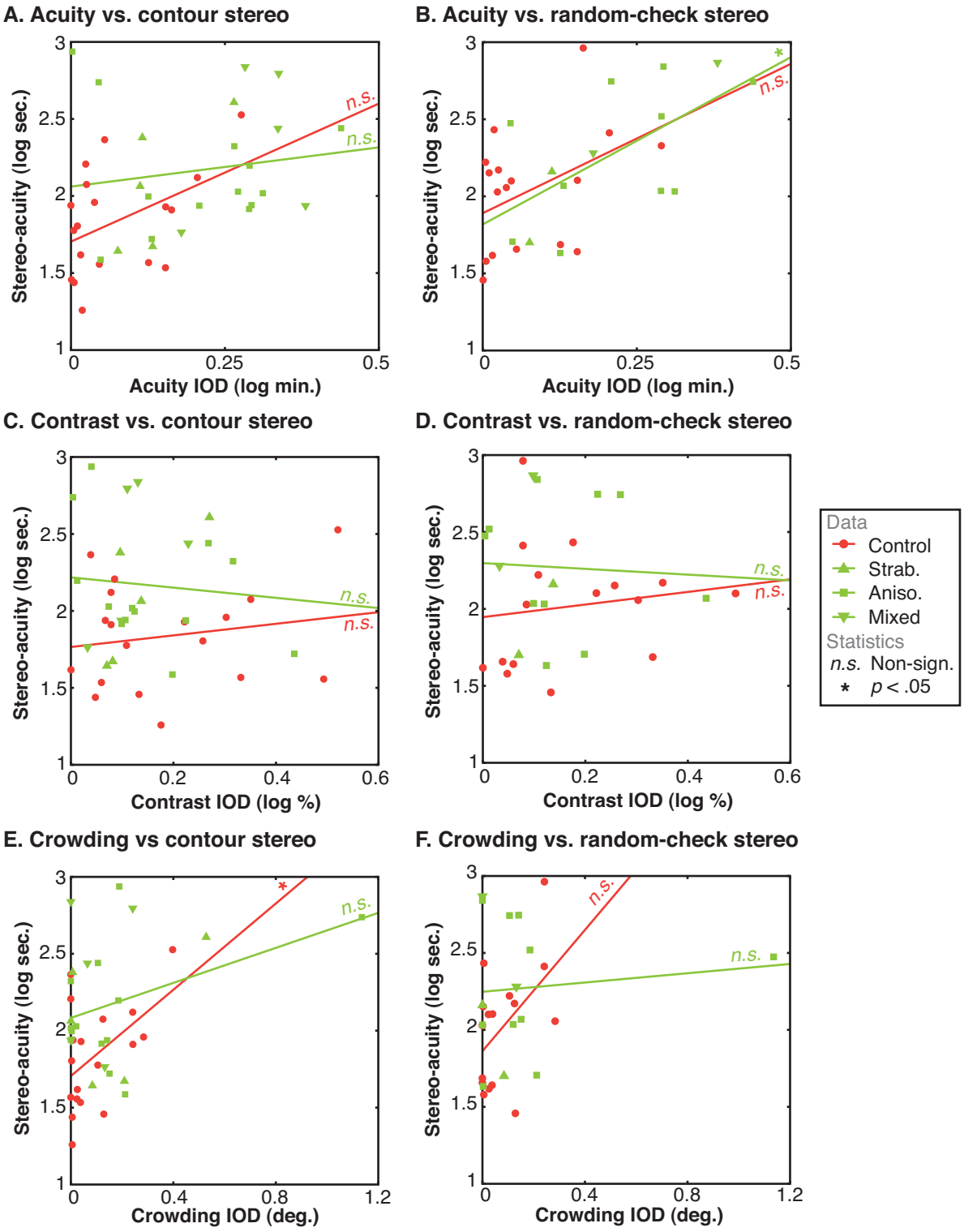
In the main-text, Figure 4 displays the stereo-acuity thresholds for those with measurable stereo-abilities. To examine the relationship between stereo-vision and interocular differences (IODs) in each of the three spatial tasks (acuity, contrast-detection, and crowding), we divided children into "stereo" and "non-stereo" categories based on their ability to achieve measureable stereo-acuity thresholds (i.e. below 1000" of arc) on *either* of the two stereo-tests. As outlined in-text, we considered this to be the conservative method of comparison rather than assigning arbitrary values to the "non-stereo" children in each group. In this section we examine the correlation between the stereo-thresholds achieved by children in the "stereo" category and their IODs in the three spatial tasks.

Because it is the existence of any interocular difference that may disrupt stereo-acuity (Goodwin & Romano, 1985), the direction of interocular differences is unimportant. We thus took the absolute interocular difference in the three spatial tasks. This does not affect data in the three clinical groups considerably, but does alter the distribution of control values. The relationship between contour-based stereo-thresholds and absolute interocular differences in acuity are plotted in Supplementary Figure 4A. Overall, this relationship is significant, with a correlation of $r(39)=0.41$, $p=0.008$. Because the selection of only those children with measureable stereo-thresholds reduces group numbers substantially (particularly for the strabismic and mixed groups), here we combine the three clinical groups into a single category. When combined in this way, there was a marginally non-significant correlation with interocular acuity differences in the control

group (*r(16)*=0.44, *p*=0.068), and a non-significant relationship for the clinical group (*r(21)*=0.15, *p*=0.51). Both groups can however be seen to follow roughly the same pattern, albeit noisily. Similarly, Supplementary Figure 4B plots the relationship between the random-check stereogram version of our task and IODs in acuity. Overall, this relationship was highly significant, with *r*(28)=0.57, *p*=0.001. When divided by category, there was a non-significant correlation for the control group (*r(14)*=0.34, *p*=0.20), though this data can be seen to follow the same trend as the clinical group for whom the relationship was significant (*r(12)*=0.63, *p*=0.016).

Supplementary Figure 4C plots the relationship between contrast-detection thresholds and contour-based stereo. Overall there is no significant correlation between the two (*r*(39)=-0.03, *p*=0.84), which is true for the individual groups as well (control: *r(16)*=0.17, *p*=0.49; clinical: *r(21)*=-0.09, *p*=0.70). The same is true for random-check stereo-acuities, as plotted in Supplementary Figure 4D. There is no significant correlation either overall (*r*(28)=0.01, *p*=0.94), or for the separate groups (control: *r(14)*=0.14, *p*=0.60; clinical: *r(12)*=0.05, *p*=0.87).

For contour-based stereo, the relationship with crowding is similar to that of acuity. The two are plotted in Supplementary Figure 4E, and overall this relationship is highly significant (*r*(39)=0.39, *p*=0.011). When broken down, this relationship is significant for the control group (*r(16)*=0.49, *p*=0.038) but not for the clinical group (*r(21)*=0.34, *p*=0.109). In contrast, there was no significant relationship between random-check stereo-acuities and crowding for the data as a whole (*r*(28)=0.23, *p*=0.22), nor for the clinical group separately (*r(12)*=0.10, *p*=0.73), though the control children show a marginally non-significant correlation (*r(14)*=0.49, *p*=0.06).



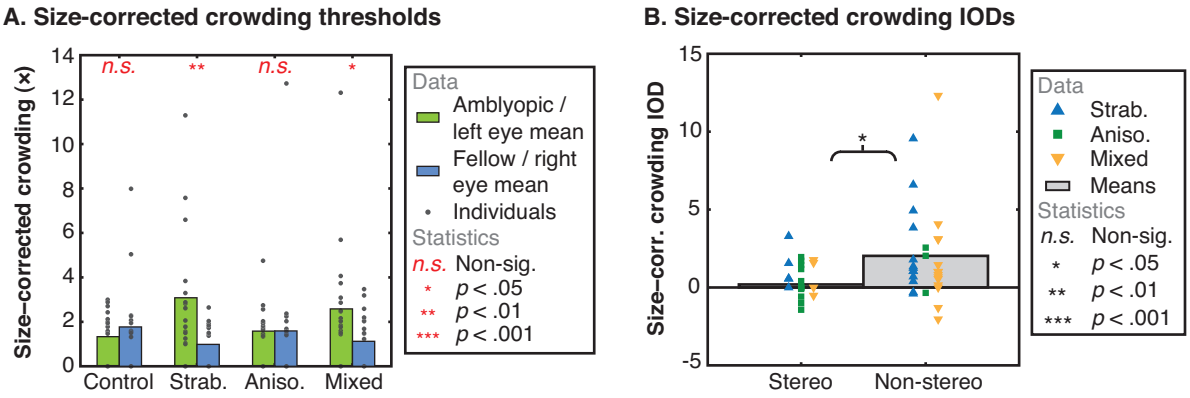
Supplementary Figure 4. Correlations between the three spatial tasks and stereo-acuity thresholds, for observers with measurable stereo-acuity. **A.** The relationship between contour-based stereo-acuity thresholds (in log seconds of arc) and interocular differences (IODs) in acuity (in log minutes). Children are divided into "control" and "clinical" categories, though separate symbols show whether clinical children are derived from the strabismic (upright triangles), anisometropic (squares), or mixed (inverted triangles) groups. The correlation for each group is shown at the right-hand side of the line of best fit. **B.** Random-check stereo-acuity thresholds, plotted against IODs for acuity, as in panel A. **C.** Contour-based stereo and IODs for contrast detection, the latter plotted in log Weber percentage units. **D.** Random-check stereo-acuity thresholds and contrast-detection IODs. **E.** IODs for crowding (in degrees of visual arc), plotted against contour-based stereo-acuity thresholds. **F.** Crowding IODs and random-check stereo-acuity thresholds.

Together, we see a similar pattern in these results as when children are divided into "stereo" and "non-stereo" categories (as in Figure 5 in-text). Namely, those with poor stereo-acuity have large interocular differences in both acuity and crowding. The correlation between IODs in acuity and stereo-thresholds has been observed previously in adults (Weakley, 2001; McKee, Levi, & Movshon, 2003; Levi, McKee, & Movshon, 2011); here we demonstrate that this is true for both contour-based and random-check stereo stimuli. Differences emerge with crowding however, where we demonstrate a strong relationship between contour-based stereo-acuity and crowding, while the relationship between random-check stereo-acuity and crowding is non-significant. This could be due to the stronger demands placed on acuity by the random-check stereograms – the individual checks in these stimuli must be resolved in order to detect the interocular correlations. The relationship with acuity may dominate these thresholds as a result, whereas the additional low spatial frequency information in the contour-based stereograms may allow the relationship with crowding to emerge.

Size-corrected crowding values

In the "normal" visual system, the extent of foveal crowding in adults has been found to follow the size of the stimuli used in testing (Levi, Klein, & Hariharan, 2002), while in the periphery (Levi, Hariharan, & Klein, 2002b; Tripathy & Cavanagh, 2002) and the amblyopic fovea (Levi, Hariharan, & Klein, 2002a), crowding is largely size-invariant. In our study, the large variation in acuity within our population necessitated that stimuli in the remaining tasks was scaled to clearly visible sizes to ensure visibility. However, this variation in stimulus size raises the possibility that the elevations in foveal crowding seen in some groups could simply reflect the variation in stimulus size (and the corresponding variation in crowding). To examine this possibility, our measured extents of crowding (in degrees of visual angle) were divided by the size of stimuli used in this task (also in degrees, which corresponds to values 2.5× acuity thresholds). The resulting values represent the spatial extent of foveal crowding expressed in multiples of stimulus size. Note that this is equivalent to expressing the extent of crowding in units of "letter widths", as is sometimes performed for crowding experiments.

Supplementary Figure 5A plots the spatial extent of this size-corrected foveal crowding in each eye for all children in our study. The same pattern of results as in Figure 2C is clearly evident. There is no interocular difference between the degree of crowding in either of the control ($t(18)$=-1.16, p=.26) or anisometropic ($t(15)$=-0.02, p=.99) groups, while a clear interocular difference is evident in the strabismic group ($t(17)$=3.38, p=.004) and a marginally significant difference exists for the mixed group ($t(18)$=2.10, p=.049). On top of this, there is still a clear degree of crowding in both eyes of controls and the fellow eyes of all clinical groups (slightly below two times the stimulus size in each case), with a combined analysis showing a significant difference ($t(90)$=5.63, p<.001). This is consistent with prior studies (Jeon et al., 2010).



Supplementary Figure 5. Size-corrected crowding values. **A.** When crowding values are expressed in multiples of acuity thresholds, the degree of foveal crowding remains significant in all eyes, as do interocular differences for the strabismic and mixed groups (data as in Figure 2 in-text). **B.** Interocular crowding differences are also larger for "non-stereo" children than for those with measurable stereo-vision (data plotted as in Figure 5 in-text).

It is also possible that the higher levels of crowding observed in our non-stereo subjects (Figure 5C of the main text) reflects the potential size-tuning of foveal crowding. To examine whether the larger IODs persisted after size correction, we split our observers into "stereo" and "non-stereo" categories, as in the main text, and plot size-corrected crowding extents in Supplementary Figure 5B. Both the pattern of individual observers, and the large group difference on average are maintained, which remains significant ($t(51)$=-2.39, p=.021). The relationship between stereo-vision and crowding is thus independent of stimulus size.

Supplementary References

Atkinson, J., Anker, S., Evans, C., & McIntyre, A. (1987). The Cambridge Crowding Cards for preschool visual acuity testing. *Transactions of the Sixth International Orthoptic Congress*, 482-486.

Atkinson, J., & Braddick, O. J. (1983). Assessment of visual acuity in infancy and early childhood. *Acta Ophthalmologica, 157 Suppl*, 18-26.

Bland, M. J., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307-310.

Ellemberg, D., Lewis, T. L., Hong Liu, C., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Research, 39*(14), 2325-2333.

Goodwin, R. T., & Romano, P. E. (1985). Stereoacuity degradation by experimental and real monocular and binocular amblyopia. *Investigative Ophthalmology & Visual Science, 26*(7), 917-923.

Jeon, S. T., Hamid, J., Maurer, D., & Lewis, T. L. (2010). Developmental changes during childhood in single-letter acuity and its crowding by surrounding contours. *Journal of Experimental Child Psychology, 107*(4), 423-437.

Levi, D. M., Hariharan, S., & Klein, S. A. (2002a). Suppressive and facilitatory spatial interactions in amblyopic vision. *Vision Research, 42*(11), 1379-1394.

Levi, D. M., Hariharan, S., & Klein, S. A. (2002b). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision, 2*, 167-177.

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

Levi, D. M., McKee, S. P., & Movshon, J. A. (2011). Visual deficits in anisometropia. *Vision Research, 51*(1), 48-57.

McKee, S. P., Levi, D. M., & Movshon, J. A. (2003). The pattern of visual deficits in amblyopia. *Journal of Vision, 3*(5), 380-405.

Ono, M. E., Rivest, J., & Ono, H. (1986). Depth perception as a function of motion parallax and absolute-distance information. *Journal of Experimental Psychology: Human Perception and Performance, 12*(3), 331-337.

Pan, Y., Tarczy-Hornoch, K., Cotter, S. A., Wen, G., Borchert, M. S., Azen, S. P., et al. (2009). Visual acuity norms in pre-school children: The multi-ethnic pediatric eye disease study. *Optometry & Vision Science, 86*(6), 607-612.

Richards, W. (1970). Stereopsis and stereoblindness. *Experimental Brain Research, 10*(4), 380-388.

Richards, W. (1971). Anomalous stereoscopic depth perception. *Journal of the Optical Society of America, 61*(3), 410-414.

Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. *Vision Research, 42*(20), 2357-2369.

Weakley, D. R. (2001). The association between nonstrabismic anisometropia, amblyopia, and subnormal binocularity. *Ophthalmology, 108*(1), 163-171.