Supplementary Material

Table A1. Clinical characteristics.

SID	Age	Gender	Type	Refractive Error	Visual Acuity (Bailey/Lovie)	Ocular Alignment (distance)	Stereo acuity (arc sec)	Fixation	History
A1	53	M	Aniso	R: -0.75/-1.00 X 100 L: -1.75/-0.75 X 70	20/20+2 20/63+2	3Δ Exophoria	40	central	
A2	46	F	Aniso	R: +0.25/-0.50 X 90 L: +3.75/-100 X 30	20/12.5-2 20/50+2	Ortho	70	central	Many studies
S1	68	F	Strab (mixed)	R: +5.50/-2.00 X 5 L: +2.25/-0.75 X 25	20/63+1 20/20-1	8ΔR. XT	Failed	RE: Unsteady	Patched and surgery
S2	33	F	Strab (mixed)	R: +6.00 L: +5.50/-1.00 X 105	20/63 20/20-2	4ΔR. XT @ Near	Failed	RE: Unsteady	Unknown
S3	34	M	Strab (mixed)	R: -1.25 L: +3.00	20/16-2 20/300	35Δ L. ET;	Failed	LE: Unsteady, 3ΔNasa	Unknown
S4	68	M	Strab (mixed)	R: -11/-1.50 X 30 L: -10/-0.75 X 140	20/80+1 20/16	12Δ R. XT@ dist; 6 Δ L. Hypo	Failed	RE: Unsteady central	Unknown
S5	56	M	Strab (mixed)	R: -1.00/-0.25 X 160 L: +0.50/-0.50 X 20	20/20 20/80+2	L. XT 4Δ@ dist; 18Δ @ Near	Failed	LE: Unsteady central	Patched at 6 Surgery at 7
S6	59	M	Strab (mixed)	R: Plano L: +2.00/-0.75 X 110	20/20+2 20/125+2	30Δ L. ET	Failed	RE: Unsteady central	None
S-NA	30	F	Strab (mixed)	R: +0.75 L: +3.00/-0.75 X 40	20/16-2 20/20-1	10ΔAlt. HT@ dist; 15 Δ @ Near	Failed	Central (both eyes)	Congenital Superior oblique palsy

ET – esotropia; XT – exotropia; HT – hypertropia; Alt – Alternating

Blue denotes anisometropic amblyopia (A), while red denotes strabismic and mixed amblyopia (S). S-NA is strabismic with no amblyopia.

Parallel vs. Intersecting Burr functions

While the Burr functions in Figs. 3 and 9 for control observers are almost completely superimposed, showing that the reaction-time difference between their two eyes is nearly null for all contrasts, we find different patterns in the amblyopic observers. In this section we ask whether delays in the amblyopic eyes RTs can be understood by shifting the Burr fit of the fellow eye (Fig. A1 and Table A2).

The Burr fit, Eq (1), has two parameters, steepness and asymptote, which affect the RT. To isolate which is responsible for the response delay in the AE, we fixed one parameter to be the NAE fit value, and varied the other parameter when fitting the AE's data. In other words, we try to interpret the AE's data by shifting the NAE's fitting curve in two different ways, intersecting (fixing the asymptote) or parallel (fixing the steepness).

We used the Akaike Information Criterion (AIC), a measure of the relative goodness of

fit of a statistical model developed by Akaike (1974), to perform a likelihood analysis on which type of shift is better supported by the data. Let K be the number of estimated parameters in the model and L_{Max} be the maximized value of the likelihood function for the model, AIC is defined as $AIC = 2K - 2\ln L_{Max}$. Assuming that the errors are normally distributed and independent, after ignoring the constant term, AIC is given by $AIC = N\ln\left(\frac{\chi^2}{N}\right) + 2K$, (2)
where χ^2 is the residual sum of square in the least squares fitting and N is the number of observed data points. To give a greater penalty for additional parameters, Burnham and Anderson (2002) recommended the AIC with a correction for finite sample sizes (AICc), which is given by,

$$AICc = AIC + \frac{2K(K+1)}{N-K-1}. \quad (3)$$

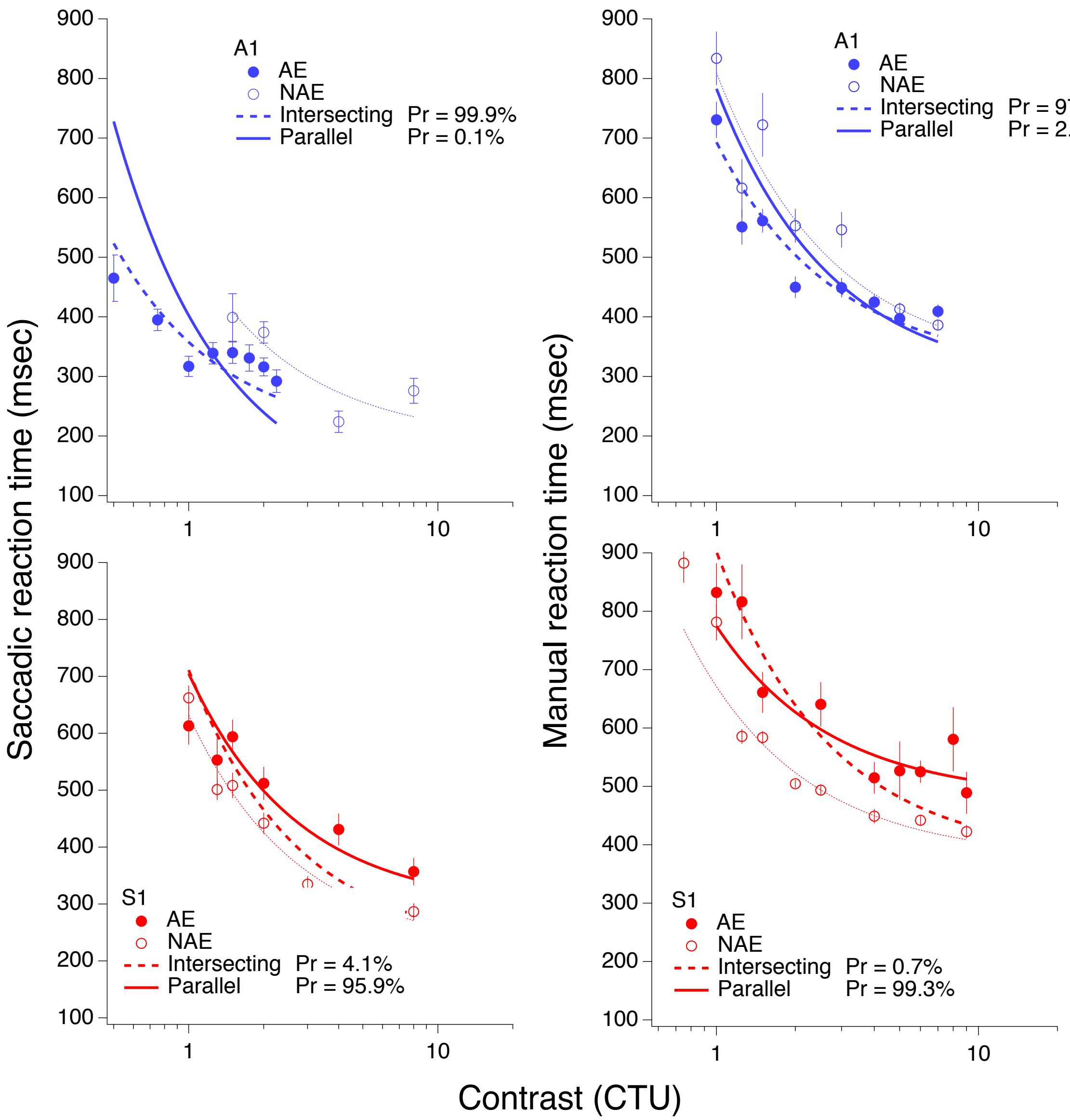

Figure A1. Saccadic (left panels) and manual (right panels) RT for two observers, A1 and S1. Thin dotted lines show the fit the NAE's data using Eq (1). The thick curves are shifted versions of the NAE fit, either intersecting (thick dashed curve) or parallel (thick solid curve), to fit the AE's data (see text for details).

We first fit the NAE's data using Eq (1), the thin dotted curves in Figure A1, then we shifted this curve, either intersecting (thick dashed curve) or parallel (thick solid curve), to fit the AE's data. When fitting the NAE's data, there are three parameters to be estimated, the steepness alpha, the asymptotic RT value, and the variance of residuals distributions, i.e., K = 3. When shifting the NAE's fitting curve to fit the AE's data, there are two parameters to be estimated (the third one is fixed), i.e., K = 2. For each observer, we used Eq (3) to calculate AICc scores of the two shifts (Table A2), i.e., AICparallel and AICintersecting, the one with the lower AICc score is more likely to be correct for interpreting the AE's response delay. The probability that the parallel shift is correct is given by

$$Prob = \frac{e^{-0.5\Delta AIC}}{1+e^{-0.5\Delta AIC}} \qquad (4)$$

where $\Delta AIC = \text{AICparallel} - \text{AICintersecting}$ (Motulsky & Christopoulos, 2004).

Figure A1 shows examples of this approach for both saccadic (left panels) and manual (right panels) RT for two observers, A1 and S1. For observer A1, for both tasks, the intersecting fit has a much higher probability of being correct. In contrast, for S1, the parallel shift has the higher probability. Table A2 shows the results for all observers and both tasks. For the saccadic task, it is clear that the parallel shift provides a better fit to the AE data of all 3 strabismic amblyopes and one of the two anisometropes (A2). The intersecting shift provides a better fit to the AE of the second anisometropic amblyope (A1), and there is no clear advantage of either for the neurotypical observer (C2) or the strabismic without amblyopia (S-NA). Similarly, for the manual task, anisometropic subject (A1), the intersecting shift provides a better fit to the AE, while for 3 of the strabismic observers, (S1, S3 and S6) the parallel shift provides a superior fit.

Table A2: AIC and Fit Probability

Saccadic RT:

Observers	AIC_inter	AIC_parallel	Prob_inter	Prob_parallel
C2	8.22	8.33	51.4%	48.6%
S-NA	9.27	8.87	45.0%	55.0%
A1	16.30	30.77	99.9%	0.1%
A2	16.05	11.46	9.1%	90.9%
S1	18.56	12.25	4.1%	95.9%
S2	29.25	27.24	26.8%	73.2%
S3	24.47	14.61	0.7%	99.3%

Manual RT:

Observers	AIC_inter	AIC_parallel	Prob_inter	Prob_parallel
C2	17.67	14.23	15.2%	84.8%
C3	11.31	12.74	67.2%	32.8%
S-NA	11.76	11.49	46.6%	53.4%
A1	17.42	24.44	97.1%	2.9%
A2	19.68	20.29	57.6%	42.4%
S1	16.09	6.19	0.7%	99.3%
S2	5.08	16.45	99.7%	0.3%
S3	18.39	12.06	4.1%	95.9%
S4	13.14	12.65	43.8%	56.2%
S5	19.13	18.88	46.9%	53.1%
S6	17.80	2.49	0.05%	99.95%

Is the parallel shift necessary and sufficient for interpreting the AE's response delay?

Although the AIC likelihood analysis can tell us which type of shift is more likely to explain the AE's response delay, it cannot answer the question of whether a shift is both necessary and sufficient. For example, the fit to one eye's response data might also provide a good fit to the other eye's data, with no shift being necessary, i.e., no shift significantly improves the fit to the other eye's data. This might be expected in the neurotypical observers, or if equating the contrast completely superimposes the data of the two eyes. On the other hand, if a parallel shift is necessary, we need to check whether it is sufficient, i.e., further changing the other model parameter (alpha) does not significantly improve the fit. In order to answer this question, we performed F-tests to compare three nested models when fitting the amblyopic or non-dominant eye's response data:

(1) NAE's fit, Eq (1) with two model parameters fixed (given by the best fit to the NAE's data);
(2) Parallel shift, Eq (1) with the parameter alpha fixed (given by the best fit to the NAE's data);
(3) Full model of Eq (1) with two free model parameters.

If Model a (a = 1 or 2) is nested within Model b (b = 2 or 3), the F-test that assesses whether Model 2 significantly improves data fitting is given by,

$$F_{a,b} = \frac{\frac{\chi^2(a)-\chi^2(b)}{\nu(a)-\nu(b)}}{\frac{\chi^2(b)}{\nu(b)}}, \qquad (5)$$

where χ^2 is the residual sum of square in the least squares fitting and ν is the number of degrees of freedom. Eq (5) compares the variance between models *a* and b with the variance inside model b and has F distribution with $[\nu(a) - \nu(b),\ \nu(b)]$ degree of freedom. When the $F_{a,b}$ value is large enough, Model *a* can be rejected at a small false-rejection probability $p(F_{a,b})$.

Table A3 shows the outcome of these tests: (1) to compare Models 1 and 2 with the $F_{1,2}$ value and $p(F_{1,2})$ in the line of Model 2; and (2) to compare Models 2 and 3 with the $F_{2,3}$ value and $p(F_{2,3})$ in the line of Model 3. We highlight (in bold) those cases where the parallel shift (Model 2) is both necessary and sufficient, i.e., Model 1 can be rejected based on $p(F_{1,2}) < 0.05$ but Model 2 cannot be rejected based on $p(F_{2,3}) > 0.05$. Consistent with F-test, for those cases where the parallel shift is both necessary and sufficient, Model 2 has the smallest AICc value, providing the best fit to the AE's response. For observer S1, the parallel shift is both necessary and sufficient for both tasks. For A1, the parallel shift is neither necessary nor sufficient for the manual task, and necessary but not sufficient for the saccadic task (because the further intercepting shift provides the better fit). Interestingly, based on AICc analysis, for the strabismic observer with no amblyopia (S-NA) the fit to the dominant eye's data (Model 1) provides the best fit (the smallest AICc) to the non-dominant eye's response.

Table A3: F-tests

Saccadic RT:

Observer:	Model	Nf	χ^2	F-test	p(F)	AICc
C2	1	7	10.19			4.60
	2	6	10.19	0.00	1	8.33
	3	5	9.74	0.23	0.95	13.58
S-NA	1	5	8.55			5.13
	2	4	6.93	0.93	0.54	8.87
	3	3	6.92	0.00	1	18.86
A1	1	7	469.07			35.24
	2	6	168.21	10.73	0.005	30.77
	3	5	7.33	109.73	0	11.30
A2	1	7	41.40			15.82
	2	6	15.05	10.50	**0.005**	**11.46**
	3	5	14.82	0.08	**0.996**	16.93
S1	1	5	48.22			15.50
	2	4	12.18	11.83	**0.016**	**12.25**
	3	3	5.46	3.70	**0.155**	17.43
S2	1	8	120.61			25.93
	2	7	95.36	1.85	0.21	27.24
	3	6	14.13	34.50	0.0002	14.86
S3	1	5	220.32			24.62
	2	4	18.04	44.84	**0.001**	**14.61**
	3	3	13.85	0.91	**0.55**	23.02

Manual RT

Observer:	Model	Nf	χ^2	F-test	p(F)	AICc
C2	1	5	59.77			16.79
	2	4	16.94	10.11	**0.02**	**14.23**
	3	3	16.79	0.03	**1**	24.17
C3	1	5	31.79			13.00
	2	4	13.21	5.62	0.06	12.74
	3	3	10.00	0.96	0.53	21.06
S-NA	1	7	15.64			8.03
	2	6	15.12	0.21	0.97	11.49
	3	5	13.93	0.42	0.84	16.44
A1	1	7	104.24			23.20
	2	6	76.29	2.20	0.18	24.44
	3	5	20.04	14.04	0.01	19.34
A2	1	7	62.79			19.15
	2	6	45.40	2.30	0.17	20.29
	3	5	41.13	0.52	0.78	25.10
S1	1	8	92.93			23.58
	2	7	9.20	63.74	**0.00**	**6.19**
	3	6	7.14	1.73	**0.26**	8.71
S2	1	8	42.37			16.51
	2	7	28.73	3.32	0.07	16.45
	3	6	4.71	30.62	0.00	4.97
S3	1	5	76.38			18.26
	2	4	11.81	21.87	**0.01**	**12.06**
	3	3	11.76	0.01	**1**	22.04
S4	1	6	17.88			9.36
	2	5	15.68	0.70	0.67	12.65
	3	4	15.47	0.05	1.00	19.55
S5	1	5	41.02			14.53
	2	4	36.78	0.46	0.79	18.88
	3	3	6.40	14.23	0.03	18.39
S6	1	7	124.44			24.62
	2	6	4.91	146.12	0.00	2.49
	3	5	2.21	6.13	0.03	1.69

The bolded values are those cases in which the parallel shift with the parameter alpha fixed (Model 2) is both necessary and sufficient for the task.

References:

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on, 19*(6), 716-723.

Burnham, K. P., & Anderson, D. (2002). *Model selection and multi-model inference: a practical information-theoretic approach (2nd ed.)*: Springer-Verlag.

Motulsky, H., & Christopoulos, A. (2004). *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*: Oxford University Press.