Vision Research 131 (2017) 106–119

Contents lists available at ScienceDirect

Vision Research

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

Disparity configuration influences depth discrimination in naïve adults, but not in children

Laurie M. Wilcox [a,*], Brittney Hartle [a], Aliya Solski [a], Kevin J. Mackenzie [a], Deborah Giaschi [b]

[a] Centre for Vision Research, Department of Psychology, York University, Toronto, Ontario M3J 1P3, Canada
[b] Department of Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, BC V6H 3V4, Canada

ARTICLE INFO

Article history:
Received 13 October 2016
Received in revised form 21 December 2016
Accepted 28 December 2016
Available online 25 January 2017

Keywords:
Stereopsis
Depth discrimination
Development
Perceptual organization

ABSTRACT

We report a series of experiments in which we assess depth discrimination performance in adults and children using a disparity-balanced target configuration to avoid the effects of anticipatory vergence eye movements. In our first study we found that children outperformed adults by a substantial margin, and the adults were consistently near chance. This was surprising given that we initially tested naïve adults to provide a benchmark for the children's data, and all observers met the criterion for stereoacuity. In subsequent experiments we recruited groups of inexperienced adult observers and assessed the role of a wide range of spatial and temporal factors in this apparent deficit. We found that the adult performance remained poor in spite of changes to the stimulus layout, exposure duration, and spatial scale. The only manipulations that improved performance were those that limited the binocular disparity to a single sign. We conclude that these data reflect a form of involuntary disparity pooling that makes it difficult for naïve observers to judge depth from disparity from multiple targets. The absence of this effect in children likely reflects the late maturation of global processes and depth cue integration.

1. Introduction

Stereopsis provides precise estimates of the position of an object relative to the distance at which the eyes are fixating (Wheatstone, 1938). For instance in their survey paper, Coutant and Westheimer (1993) report that 97% of their inexperienced observers could discriminate depth from disparities as small as 2.3 arc min. Even without extensive practice, 80% of their group were able to see depth from 30 arc sec of disparity. In psychophysical studies of stereopsis, experienced observers often have stereoscopic thresholds in the range of 8 arc sec (see McKee, 1983). Given such precision, it is not surprising that much of the research about human stereopsis has focused on the smallest disparity that can be reliably discriminated: stereoacuity. However, stereopsis also provides reliable depth information for extremely large depth differences that cannot be fused into a single scene (Foley, Applebaum, & Richards, 1975; Mitchell, 1969; Ogle, 1952, 1953; Tschermak & Hoefer, 1903; Wilcox & Hess, 1995). Instead these large disparities give rise to double vision, or diplopia, but still convey depth information. Depth processing of disparities beyond the fusion limit is referred to as coarse stereopsis (reviewed in Wilcox & Allison, 2009).

We recently reported that coarse stereopsis matures to adult levels before the age of 4 years, and that fine stereopsis (disparities in the fused range), at least for small disparities, continues to mature until the age of 14 years (Giaschi, Lo, Narasimhan, Lyons, & Wilcox, 2013). The latter finding is consistent with previous research that used commercial tests such as the Titmus, Randot, Frisby, or TNO to study the maturation of stereoacuity (Ciner, Schieman, Schanelklitsch, & Weil, 1989; Cooper, Feldman, & Medlin, 1979; Fox, Patterson, & Francis, 1986; Heron, Dholakia, Collins, & McLaughlan, 1985; Leat, St Pierre, Hassan-Abadi, & Faubert, 2001; Romano, Romano, & Puklin, 1975; Simons, 1981; Tomac & Altay, 2000). The earlier maturation of coarse stereopsis may explain why it is often spared when fine stereopsis is disrupted by amblyopia (Giaschi, Narasimhan, Solski, Harrison, & Wilcox, 2013). Taken together, our work on the typical and atypical development of fine and coarse stereopsis supports previous adult studies which have suggested that different neural mechanisms mediate fine and coarse stereopsis (Hess & Wilcox, 1994, 2008; Kovács & Fehér, 1997; Langley, Fleet, & Hibbard, 1999; McKee, Verghese, & Farell, 2004, 2005; Schor, Edwards, & Sato, 2001; Wilcox & Hess, 1995, 1996, 1997).

---

[*] Corresponding author at: 4700 Keele St., B0003H Lassonde, York University, Toronto, Ontario M3J 1P3, Canada.

E-mail addresses: lwilcox@yorku.ca (L.M. Wilcox), brit1317@yorku.ca (B. Hartle), asolski@hotmail.com (A. Solski), Kevin.MacKenzie@oculus.com (K.J. Mackenzie), giaschi@mail.ubc.ca (D. Giaschi).

http://dx.doi.org/10.1016/j.visres.2016.12.010
0042-6989/

In our previous developmental studies (Giaschi, Lo et al., 2013; Giaschi, Narasimhan et al., 2013), we assessed the largest disparity that could be fused for each observer, then measured accuracy for depth order judgements (“near” or “far”) for fused and diplopic cartoon characters. A single character surrounded by a zero disparity reference frame was presented centrally for 320 ms at one of a range of crossed and uncrossed disparities. While this paradigm allowed us to examine the development of the full range of stereoscopic depth perception, not just the lower limits on performance studied by other groups, it is possible that the results were influenced by vergence eye movements. Although observers were instructed to maintain fixation in the plane of the reference frame, the stimulus duration was long enough for a disparity-driven vergence signal to be initiated. The direction of the depth offset could, in theory, be determined by monitoring the vergence signal even if an eye movement was not completed because the amplitude of disparity evoked vergence responses scales with the size of the disparity offset (Busettini, Fitzgibbon, & Miles, 2001; Jones, 1980; Westheimer & Mitchell, 1969). Furthermore, we have shown recently that cartoon characters presented at large depth offsets drive vergence very effectively in young children (Meier, Giaschi, Wilcox, Seemiller, & Candy, 2016).

The initial goal of the current study was to compare depth discrimination in children and adults using a dual-target paradigm in which two stimuli were offset by equal amounts in opposite directions in depth (Edwards, Pope, & Schor, 1999; Wilcox & Hess, 1995, 1996; Ziegler & Hess, 1997). The task was to indicate whether the top or bottom cartoon character was closer to the observer; the direction of the depth offset was randomized across trials. With this stimulus configuration, the disparity-driven vergence response is effectively nulled because disparity is ‘balanced’ about the fixation plane.

To anticipate the results, we did find in Experiment 1 that 4–6 year-olds perform well across the fine and coarse disparity ranges, which replicates our previous results (Giaschi, Lo et al., 2013). We were surprised to find, however, that naïve adult observers with normal stereoacuity performed close to chance at all disparities. This finding led to a series of experiments with the common goal of identifying the stimulus properties that contributed to the difficulty that the adult observers exhibited with the dual-target depth discrimination task.

To summarize, we show that the 2D stimulus configuration does not influence performance on this task, nor does exposure duration or spatial distribution of attention. In fact, the only manipulations that improved adult performance were those that limited the binocular disparity to a single sign (crossed or uncrossed), as in the single target stimulus used by Giaschi, Lo et al. (2013). These results are consistent with a form of involuntary perceptual grouping of binocular disparity that makes it difficult for naïve adult observers to report depth sign when objects are presented at crossed and uncrossed disparities simultaneously. The fact that children were impervious to this disruptive influence is discussed in the context of the development of 2D and 3D depth cue integration.

2. Experiment 1 – Suprathreshold depth discrimination in adults and children

As outlined above, to help our observers maintain convergence on the central fixation stimulus and to avoid any influence of eye movements, we offset two identical targets in equal and opposite directions in depth. Depth percepts were assessed for the same range of suprathreshold disparities tested by Giaschi, Lo et al. (2013) both within and beyond Panama’s fusional area for both young (4–6 years) and adult observers.

2.1. Methods

2.1.1. Participants

Eighteen children (aged 4–6 years) were recruited from University of British Columbia daycare centres. Informed assent was obtained prior to testing from all children, and written informed consent was provided by a parent. Children were given stickers and small games for their participation. Seventeen adults (aged 18–33 years) were recruited via York University’s undergraduate participant pool. These observers received course credit for their participation in the study. All participants were screened for visual deficits using the Randot Pre-school stereotest (Stereo Optical Co.) and the Regan high-contrast letter chart (Regan, 1988) at the beginning of the test session. All participants included in the data analysis achieved a best-corrected visual acuity of at least 20/25 (Dobson, Clifford-Donaldson, Green, Miller, & Harvey, 2009) on the Regan chart, and a stereoacuity of at least 60arcsec on the stereotest (Birch et al., 2008). One child failed to complete the testing and two adults were excluded: one did not meet the inclusion criterion for stereopsis and the other did not complete testing. The remaining participants passed the vision screening and had no known eye or vision disorders. All test protocols used here adhered to the tenets of the Declaration of Helsinki. The young observers were tested in their daycare, while the adults were tested at York University.

2.1.2. Stimuli and apparatus

Our stimuli were the same nine greyscale characters used by Giaschi, Lo et al. (2013) and Giaschi, Narasimhan et al. (2013) as illustrated in Fig. 1. The stimuli were displayed on a CRT monitor (1024 × 768) with a refresh rate of 120 Hz. The display had a background luminance of 140 cd/m^2 and was positioned 100 cm from the observer. Note that although the height of the individual characters varied slightly (1.8–2.1 deg) the horizontal extent was fixed at 1.9 deg. The fixation marker was the outline of a happy face with a diameter of 1.5 deg. The vertical edge-to-edge separation between each character and the fixation marker was 1.9 deg. Two identical characters were presented to each eye, one above and one below the fixation marker. The total vertical extent of the stimulus configuration was 9.0–9.7 deg. Stereoscopic depth was achieved using Crystal Eyes LCD shutter glasses synchronized with the display using an infrared emitter. The characters were offset in depth by equal horizontal disparities, in opposite directions. Two ranges of disparity were tested in separate blocks; five fine (0.02, 0.08, 0.17, 0.33, and 0.67 deg) and five coarse (1, 2, 2.5, 3,

Fig. 1. The set of nine characters (modified versions of Pokémon™ figures) used in the experiments described here, not drawn to scale. Figures were resized so that the width was constant. On each trial the same character was presented stereoscopically above and below fixation (see details in the text).

and 3.5 deg). In previous studies we established by measuring diplopia thresholds that the coarse disparities are in the diplopic range for both children and adults (see Giaschi, Lo et al., 2013; Giaschi, Narasimhan et al., 2013). There was no observable cross-talk or visible ghosting between the images presented to each eye. Responses were made using a gamepad.

2.1.3. Procedure

Participants first viewed the fixation marker which appeared in the centre of the display at the screen plane, followed by the test stimuli. During the test interval, two identical characters appeared above and below fixation. To accommodate our young observers, we used a relatively long exposure duration of 750 ms. On each trial the observer was asked to indicate which character appeared to be in front of the fixation marker. Prior to testing, each participant was given practice, with feedback, to familiarize them with the task. Both adults and children initially viewed a PowerPoint slideshow depicting a detective searching for cartoon characters. They then completed a set of 20 trials, with auditory feedback, in which the stimulus was presented for 1500 ms. Following training, a single test session consisted of 12 trials per disparity, with no feedback, presented across 6 blocks of trials, with the fine and coarse range tested in separate blocks. We recorded the percentage correct responses for each disparity.

2.2. Results and discussion

The results depicted in Fig. 2 replicate those of Giaschi, Lo et al. (2013) in showing that children as young as 4 years perform well on this depth discrimination task for disparities at or above 0.08 deg. Here we do not see the drop in performance at the lowest disparity reported in our earlier study, but this likely reflects the fact that the total disparity is double that used when only one target was offset in depth. Importantly, we find that the children have higher accuracy (mean = 89%) than adults, who are near chance at all test disparities (55%). In our previous study (Giaschi, Lo et al., 2013), there was a tendency for children older than 5 years to be slightly more accurate than adults at disparities between 0.17 and 1.0 deg, but this was not statistically significant.

A mixed-design analysis of variance confirmed that the difference between the adult and child performance is highly significant, $F(1,30) = 228.56$, $p < 0.001$; $\eta^2 = 0.76$. A Greenhouse-Geisser correction was applied to the factor Disparity and interaction effect to correct for the violation of Mauchly's Test of Sphericity ($p < 0.001$). There was no main effect of Disparity, $F(4.2,124.5) = 0.90$, $p = 0.47$; $\eta^2 = 0.02$, nor was there an interaction between the performance of the children and adults as a function of Disparity, $F(4.2,124.5) = 2.06$, $p = 0.09$; $\eta^2 = 0.04$. This is evident in Fig. 2; there is a large difference in performance between adults and children, but little impact of disparity.

To verify that the marked discrepancy in performance seen here was not due to a trivial difference in the experimental set up in the two locations, a group of ten children (ages 5–7) were tested at a daycare centre in Toronto (Howard Park Daycare) using the same equipment that was used to test the adult observers in Experiment 1. The results for this group of children were very similar to those obtained at UBC, with average performance at 85%, again well above that of the adult observers. We also tested a group of adult observers at UBC, and their results were virtually identical to those of the adults tested at York University. Spontaneous feedback from the young observers at both locations was very different from that obtained from the adults. Whereas the children said that they enjoyed the task and found it easy, the adults complained that it was difficult and tiring. Note that neither group of observers (child or adult) had prior experience with psychophysical testing using stereoscopic displays. Given that we anticipated that adult performance on this task would provide a benchmark for children's performance (and not vice versa), these results were puzzling. There is little precedent in the literature for children exhibiting better stereoscopic performance than adults. One exception is a study by Dowd, Clifton, Anderson, and Eichelman (1980) who found that at large disparities, children were faster and more accurate in identifying depth in random-dot stereograms. More recently, Nardini, Bedford, and Mareschal (2010) showed that children (6 years) performed better than adults on a slant discrimination task under cue conflict conditions. They reasoned that mandatory sensory fusion in adults degraded their performance when slant information provided via binocular disparity conflicted with that provided by texture. However, children were able to access these cues independently thereby avoiding the disruptive effects of conflict. In Nardini et al.'s (2010) study, adults performed better than children when depth cues were consistent, suggesting that they were able to capitalize on reduced uncertainty. Given that we used isolated stereoscopic figures and did not explicitly introduce cue conflicts, it is difficult to interpret our results in the context of these studies.

However, Harris, Chopin, Zeiner, and Hibbard (2011) reported poor adult performance in naïve observers on a task that is similar to the one used here. Harris and colleagues designed their experiment to evaluate the extent to which naïve observers were able to capitalize on viewing geometry to extract binocular disparity. They used isolated dot stimuli, positioned above and below a fixation

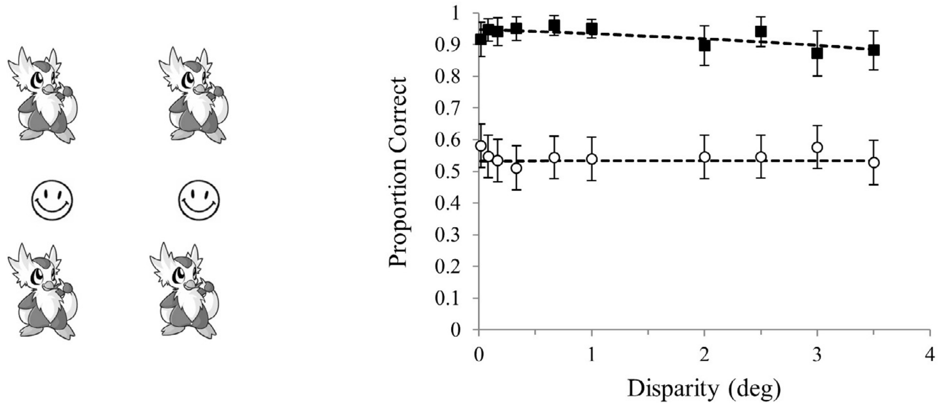

Fig. 2. The cartoon stimuli to the left are shown here as a stereogram; when the eyes are crossed to fuse the central fixation marker (happy face) the upper character will appear near, and the lower character far. Divergent fusion will reverse the depth order. The graph to the right depicts accuracy as a function of test disparity for adults (open circles) and children (filled squares). The error bars indicate the standard error of the mean.

point, offset in equal, but opposite directions in depth. In a 2IFC paradigm, observers were asked to indicate which interval contained the larger separation in depth between the two targets; the relative depth order of the two targets was either the same (top-near, bottom-near) or different (top-near, bottom-far). They recruited 24 naïve adult observers (18–38 years), *only 8 of whom could reliably perform this task*, despite the fact that they were all pre-screened for stereo-deficits, and that the test disparities were suprathreshold. Harris et al. (2011) evaluated two potential sources of error which could influence depth estimates in the observers who could perform the task – the presumed slant of the vertical horopter and rotational eye movements (see Experiment 2a below for more detail) – but concluded that neither factor mediated the biases they observed. With reference to the large number of observers who did not meet inclusion criteria, they suggested that naïve observers simply were not able to make accurate metric depth judgements based solely on binocular disparity (and eye position). Similarly, we found in Experiment 1 that naïve adults' depth discrimination performance was degraded for stimulus offset about the fixation plane. However, the fact that children performed our two-target task so well, and that naïve adults exhibit high accuracy in a single target paradigm (Giaschi, Lo et al., 2013), suggests that some other factor may be responsible for the degraded adult performance shown in Experiment 1, and potentially for the high exclusion rate in Harris et al. (2011).

In Experiment 2, we evaluated a number of configural factors that may have introduced biases in perceived depth that degraded naive adult performance. We assessed the effect of: a) the slant of the vertical horopter; b) exposure duration; c) size constancy; d) 2D perceptual grouping; and e) scale of configuration. For all these experiments we continued to use the broadband, greyscale, characters to permit direct comparison with the results obtained with adults in Experiment 1.

3. Experiment 2 – Role of configural factors in adult depth discrimination

3.1. General Methods

3.1.1. Participants

In Experiments 2a-e adult participants (aged 17–26) were undergraduate students recruited from York University via the Department of Psychology participant pool. For each study we recruited a new group of observers with a history of good ocular health, corrected to normal visual acuity and normal stereoacuity. A few of these inexperienced observers participated in up to three experiments, but in all cases we verified that there were no practice effects. Stereopsis was assessed for all participants using the Randot Preschool Stereotest and an inclusion criterion of stereoacuity less than 60arcsec was applied (observers rarely failed to meet this criterion, on average we excluded one observer per study for this reason).

3.1.2. Apparatus and stimuli

The apparatus is the same as described in Experiment 1 above. The basic stimulus configuration is the same as described in Experiment 1, with modifications outlined in each study. In Experiment 1 our adult observers found the task difficult at 750 ms viewing time. Most also reported that even during the practice session, when they were given 1500 ms to view the stimulus, they could not perform the task reliably, as indicated by the feedback. Given this, in subsequent experiments we did not increase the exposure duration further. In addition, in Experiment 1 the fixation marker disappeared when the stimuli were presented. In a follow-up study observers reported that the task was more comfortable when the fixation marker was visible throughout the trial, so in all subsequent experiments the 'happy face' remained on the screen throughout testing. Note that this procedural change did not impact performance.

3.1.3. Analysis

For each observer we calculated the proportion correct for each disparity in each test condition. Because this series of experiments was conducted over several years, different observers were tested, and there were some minor fluctuations in the test conditions (e.g. in some cases a smaller disparity range was tested or a different number of observers were involved). To permit statistical comparison of the original adult data with subsequent experiments we fit the raw data with a mixed-effect logistic regression using a logit (binomial) link function. The model was fit using the glmer() function from the lme4 package in R, which fits a generalized linear mixed model (GLMM) incorporating both fixed and random effects parameters using maximum likelihood estimation with a Laplace approximation (Bates, Maechler, Bolker, & Walker, 2015). Reported effect sizes were converted from log odds ratios into Cohen's standardized mean difference (d) values using the transformations proposed by Borenstein, Hedges, Higgins, and Rothstein (2009). To give strong control of the family-wise error rate all p-values were corrected using the Holm-Bonferroni method (Holm, 1979). The fixed effects in this model consist of our dichotomous dependent variable (accuracy) in terms of our two independent variables of Condition and Disparity. Accuracy is recorded as either 1 for a correct response, or 0 for an incorrect response, and is represented as the proportion correct within each condition. Condition is a categorical variable that contains the experimental conditions examined in each comparison. Disparity is a continuous variable that consists of all the levels of disparity tested in each experimental condition. The random effects model can handle repeated measures data by allowing within-group errors to be correlated. The random model included by-subject intercepts to take into account interobserver differences. To account for the within-subject factor of Disparity, random by-disparity slopes and intercepts were also included as random effects. The random effect of Disparity explicitly models the individual variability in the slope and intercept as a function of disparity and the correlation of the variance within-subjects.

4. Experiment 2a – Role of the slanted vertical horopter

A potential explanation for the relatively poor adult performance shown in Fig. 1 is related to the vertical alignment of the two targets. It has been reported that the vertical horopter has a top-back slant of approximately 2.1 deg (Cooper, Burge, & Banks, 2011; Tyler & Scott, 1979). This is referred to as 'Helmholtz shear' (Helmholtz, 1962) and is thought to be due to the shear of vertical corresponding points on the retina (see Howard, 2012). To summarize, if two points are positioned above and below a fixation point at zero disparity, the slanted vertical horopter will introduce a small uncrossed disparity above fixation, and a small crossed disparity below fixation. There is some suggestion in the literature that the slanted vertical horopter influences depth discrimination by introducing a 'top-back' bias (Breitmeyer, Julesz, & Kropfl, 1975; Helmholtz, 1867). As noted in Experiment 1, Harris et al. (2011) designed their stimulus configuration specifically to assess the impact of this bias on suprathreshold depth interval judgements. They confirmed that biases in their sample of 8 observers who *could* perform their depth interval judgement were not attributable to the presumed slant of the vertical horopter. Understandably, they did not evaluate whether or not this bias was responsible for the poor performance of the excluded observers. It has been proposed that the slant of the vertical horopter results from our visual experience with the natural environment (Helmholtz, 1867; Krekling & Blika, 1983). If so, it is possible that this bias would not impact children's depth judgements as strongly as those of adults, and could explain the results seen in Experiment 1. Therefore, we evaluated this possibility by replicating our original study, but with the stimuli aligned horizontally instead of vertically.

4.1. Methods

4.1.1. Participants

Sixteen observers were recruited as described in the General Methods section. One observer was excluded from the analysis due to failure to complete testing.

4.1.2. Apparatus & procedure

The apparatus and procedure were also the same, but to evaluate the effect of the slanted horopter, we repositioned the cartoon stimuli to lie in the horizontal midline of the display, to the left and right of fixation (Fig. 3 left). In this configuration, the vertical stimulus position is constant and so there should be no influence of the slanted horopter on relative depth judgements.

Testing proceeded as described in Experiment 1 for 10 disparities (0.02–3.5 deg), with an exposure duration of 750 ms, and a total of 60 trials per disparity (completed in 3 blocks of 20 trials per disparity).

4.2. Results and discussion

Fig. 3 (right) shows depth discrimination accuracy for 15 observers who viewed the horizontally aligned stimuli, compared with the results from the original 17 observers who viewed the vertically aligned patterns. The data are virtually indistinguishable, with an average near 55 percent for both groups. The results of the mixed-effect logistic regression confirmed the absence of a significant difference between the two conditions ($\beta = 0.09$, SE = 0.08, p = 0.48, d = 0.05), a lack of an effect of disparity ($\beta = -0.01$, SE = 0.03, p = 0.79, d = 0.01), and an absence of any interaction effects ($\beta = 0.06$, SE = 0.04, p = 0.33, d = 0.04).

The results of Experiment 2a are remarkably similar to Experiment 1, which supports the consistency of this deficit in naïve adult observers. If we assume that the poor adult performance shown here has the same origin as the difficulties Harris et al.'s (2011) observers had with their task, then our results also support their conclusion that biases in depth perception in their study do not reflect the slant of the vertical horopter.

5. Experiment 2b – Exposure duration

Giaschi, Lo et al. (2013), Giaschi, Narasimhan et al. (2013) presented their stereoscopic stimuli for 320 ms, but in Experiment 1 we used a longer exposure duration (750 ms) because we assumed that the addition of the second stimulus and the balanced disparity configuration might make the task difficult for young observers. Harris et al. (2011) also used a long exposure duration for each presentation interval (2000 ms per interval). It is possible that when viewing the disparity-balanced configuration in Experiment 1, adults attempted to converge off the fixation plane to reduce the amount of retinal disparity. This strategy would not help, and in fact, could be responsible for the degraded performance. That is, if observers try to maintain an off-fixation vergence state in one direction in depth, the amount of relative disparity in the stimulus positioned in the other direction would be increased. For example, at the largest test disparity (3.5 deg), fixation at the plane of the near target will result in a 7 deg disparity between this point and the far target (and vice versa). We think this strategy is unlikely because observers confirmed that they maintained fixation as requested, but also because we randomly varied which of the two stimuli was offset in the crossed/uncrossed direction. However, the 750 ms exposure duration used in Experiment 1 is sufficient to permit completion of a vergence eye movement, at least for adult observers (Rashbass & Westheimer, 1961). However, Yang, Bucci, and Kapoula (2002) showed that children take longer than adults to initiate disparity-dependent vergence eye movements. Importantly, in the 4–6 year age range tested in Experiment 1, the latencies to initiate vergence eye movements reported by Yang et al. (2002, Fig. 3) were close to 400 ms, almost twice that of the adult observers. Moreover, there is evidence that children make fewer vergence eye movements than adults when viewing stereograms (Dowd et al., 1980). To reduce the likelihood that adult's performance on our task was degraded by vergence eye movements we replicated Experiment 1 using the same stimulus configuration, but reduced the exposure duration to 150 ms (Rashbass & Westheimer, 1961).

5.1. Methods

Using the methods outlined in the General Methods section we tested a group of 17 naïve observers with normal, or corrected-to-normal, vision. All observers practiced the task using a 750 ms exposure duration; in subsequent testing the exposure duration was reduced to 150 ms. The range of test disparities was the same

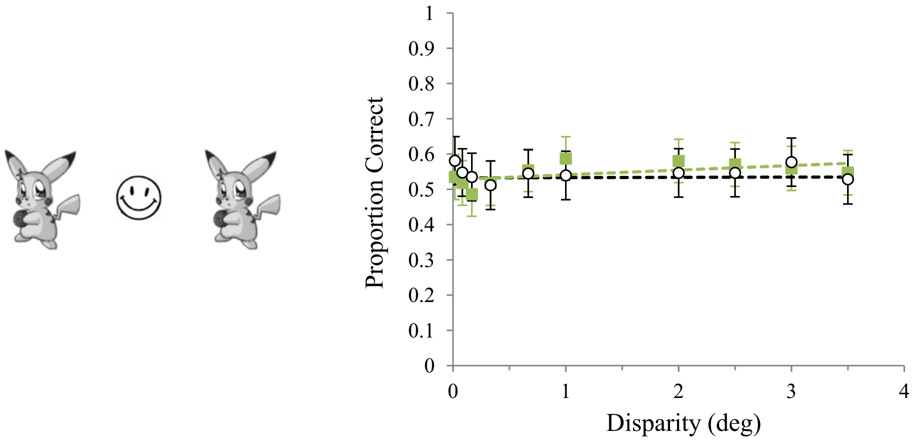

Fig. 3. One eye's view of the stimulus arrangement for Experiment 2a is shown here on the left. The depth discrimination results are shown in the graph to the right. On this, and all subsequent graphs the adult data from Experiment 1 are depicted as open circles, and the comparison data (in this case for horizontally aligned stimuli) as filled symbols. The dashed lines indicate the average marginal probabilities predicted from the logistic regression model. Error bars indicate one standard error of the mean.

as that used in Experiment 2a, but given the lack of effect of disparity, one disparity was excluded (2.5 deg) to reduce the total duration of the experiment. As in the previous experiments observers completed 20 trials per disparity in a block and completed three such blocks of trials for a total of 60 trials per disparity.

5.2. Results and discussion

The data collected with an exposure duration of 150 ms shown in Fig. 4 are similar to those obtained previously using an exposure duration of 750 ms. Three observers in this study completely inverted their responses. Unfortunately, we were not able to retest these individuals to determine if this was due to simply reversing the response options, or if they saw reversed depth, so we excluded them from the analysis (N = 14). As a result, the variance is somewhat higher. Analysis of these data using the mixed-effects logistic regression shows that there is no significant difference in accuracy between the 150 ms and 750 ms exposure durations ($\beta = 0.26$, SE = 0.16, p = 0.33, d = 0.14), nor is there a significant effect of disparity ($\beta = 0.09$, SE = 0.07, p = 0.36, d = 0.05) or interaction between exposure duration and disparity ($\beta = -0.08$, SE = 0.08, p = 0.36, d = 0.05).

We postulated that the long exposure duration used in Experiment 1 may have permitted eye movements which disadvantaged adult participants in this study. Reducing the exposure duration from 750 to 150 ms did not significantly impact accuracy. While the apparent shift in average performance was in the predicted direction, it is clear from Fig. 4 that even so, adult performance does not approach that of the young observers in Experiment 1. In a subsequent study, with a different group of observers, we assessed adult performance on our depth discrimination task with a viewing time of 320 ms. As expected, performance was statistically indistinguishable from that seen in Experiment 1 at 750 ms. In later experiments we adopted the 320 ms test duration (750 ms during practice trials) because our adult observers stated that the pacing of the trials was more comfortable.

6. Experiment 2c – Disparity-induced size constancy

In the stimulus configuration used in Experiment 1, it is possible that size constancy mechanisms altered the apparent size of the targets. This may cause the target presented with uncrossed disparity (the more distant target) to look larger than the target presented with crossed disparity. Since the same character is presented above and below fixation, this apparent size difference could be mistakenly interpreted in terms of depth. Thus, the 'larger' target would be perceived as 'near' and the 'smaller' target as 'far' – the opposite of the depth order signaled by binocular disparity. This cue conflict might interfere with depth judgments from disparity and degrade performance. This is a potential problem in all psychophysical studies of stereopsis using computerized displays, but particularly for the suprathreshold disparity range studied in our set of experiments. There is evidence that size constancy mechanisms are immature in children (Zeigler & Leibowitz, 1957). Furthermore, as outlined above, Nardini et al. (2010) have shown that at 6 years of age, children do not exhibit disrupted slant judgements when stereopsis and texture based depth cues are in conflict. This suggests that they process depth cues separately at this age, which, if also applied to size-based depth information, supports a cue-conflict based explanation for the child–adult difference in depth discrimination seen in our studies.

6.1. Methods

Eleven observers were recruited as described in the General Methods section. The apparatus and procedure were also the same, but to ensure that observers could not rely on the conflicting relative size cue, we pseudo-randomly varied the relative height of the upper and lower stimuli from trial to trial. Observers were told that they should ignore the relative size of the characters because it would be varied, and instead to attend to their displacement in depth. To avoid changes in performance due to the size disparity correlation (Tyler, 1973; Smallman & MacLeod, 1994), we kept the width fixed at 1.9 deg (as in Experiment 1), and varied the height by up to 0.2 deg. Thus, on a given trial, the same character was presented above and below the fixation marker, but its height was substantially different. Pilot trials were used to ensure that these variations were visible, and at the scale of the changes expected from size constancy. All other stimulus parameters were the same as described in the General Methods, including the edge-to-edge separation of the stimuli from fixation (1.9 deg). The exposure duration was 320 ms.

In this condition, given that the relative size information is explicitly useless, observers were compelled to rely on binocular disparity to make their depth judgements. If they could use disparity, then performance should improve relative to the original group of naïve adults. Testing proceeded as described in the General Methods section, with 20 trials per block and a total of 60 trials for each of 9 (0.02–3.5 deg) disparities.

6.2. Results

The data in Fig. 5 show depth discrimination accuracy for the stimuli that vary in size compared with fixed size stimuli

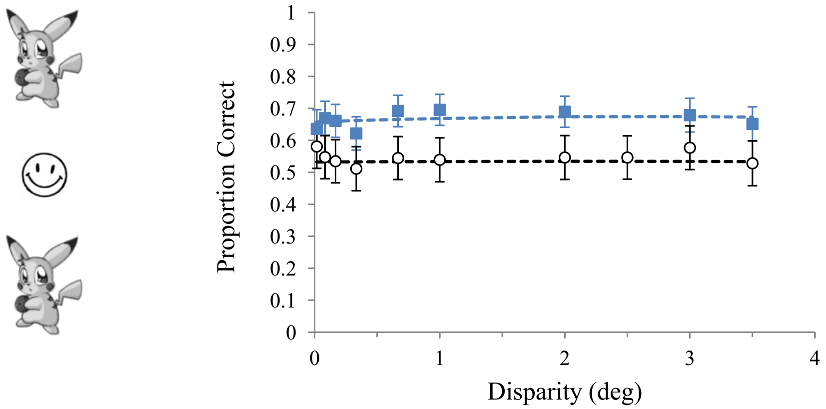

Fig. 4. The stimulus arrangement for Experiment 2b is shown here (one eye's view only) on the left. The depth discrimination results are shown in the graph to the right. Data are shown for exposure durations of 150 ms (closed squares) and 750 ms from Experiment 1 (open circles). The dashed lines indicate the average marginal probabilities across the range of test disparities predicted from the logistic regression model, and error bars represent one standard error of the mean.

presented at the same exposure duration. It is clear that the variation in size (and explicit instructions to ignore the relative size information) did not affect performance. The mixed-effect logistic regression confirmed that there was no impact of changing the relative size of the upper and lower cartoon characters on accuracy ($\beta = -0.37$, SE = 0.17, $p = 0.09$, $d = 0.20$). In addition, there was no significant effect of disparity ($\beta = 0.01$, SE = 0.05, $p = 0.96$, $d = 0.01$), nor was the two-way interaction between disparity and the size manipulation significant ($\beta = -0.06$, SE = 0.08, $p = 0.96$, $d = 0.03$).

7. Experiment 2d – 2D figural grouping

The stimuli used in Experiments 2a-c contained a number of 2D Gestalt grouping cues (Koffka, 1935). For instance, to avoid effects due to character preference in children, in a given trial the same cartoon figure was presented above and below fixation. Further, the two stimuli were always aligned vertically (or horizontally in Experiment 2a). It is possible that the alignment and similarity of the stimuli may have generated configural grouping which made it difficult for adult observers to discern the disparity sign. To reduce or eliminate their potential influence, in this study we (i) introduced a random lateral displacement to both the upper and the lower character so they were no longer vertically aligned, and (ii) changed the appearance of one of the two stimuli by scrambling the phase components of its Fourier spectrum leaving the frequency composition unaltered. We propose that if there is a disruptive 2D grouping effect that degrades depth judgements in Experiment 1, then adult performance should improve with these two manipulations. In contrast, there is evidence that children are not subject to the same 2D grouping cues (Kovács, 2000).

7.1. Methods

Participant recruitment, criteria, and screening procedures were as described in the General Methods. Of the 15 subjects in this experiment two had participated in at least one of our previous experiments; neither of these observers showed any improvement in performance that could be attributed to this limited practice. All aspects of the apparatus in this experiment are identical to those outlined in the General Methods section. The stimulus configuration in this experiment was similar to the original stimulus configuration, that is, two characters, one above and one below fixation, were displaced in opposite directions of depth by equal amounts. However, the stimuli were randomly displaced laterally, so they never appeared perfectly aligned. In one condition, the lateral offset ranged from 0.2 to 1 deg. In a second condition, we manipulated the similarity of the upper and lower targets by phase scrambling one of the two stimuli (see Fig. 6). Using the *fft* and *ifft* functions in Matlab^{tm} we randomized the phase information while leaving the amplitude spectrum unaffected. As usual on each trial, the specific cartoon character was randomly selected and the original version was presented in the upper or lower position (at random) and its phase-scrambled version was presented in the other position. As in previous experiments, we tested 10 disparity values ranging from 0.02 to 3.5 deg with 32 trials per disparity (320 trials in total). In this study, the lateral offset and phase-scrambled conditions were completed by all observers so, given the lack of effect of exposure duration shown in Experiment 2b, to reduce the duration of the study the exposure duration was reduced to 150 ms.

7.2. Results and discussion

It is evident from Fig. 6 that there was no improvement in performance due to phase scrambling or laterally offsetting the stimuli. This was confirmed by the results of the mixed-effect logistic regression that revealed a lack of significant difference in accuracy in the lateral offset condition ($\beta = -0.24$, SE = 0.10, $p = 0.06$, $d = 0.13$) relative to the two-target condition with no effect of disparity ($\beta = -0.01$, SE = 0.05, $p = 0.86$, $d = 0.01$) or interaction effects ($\beta = 0.14$, SE = 0.06, $p = 0.06$, $d = 0.07$). Similar results are seen in the comparison of the phase scrambled condition to the original two-target condition ($\beta = 0.27$, SE = 0.11, $p = 0.06$, $d = 0.15$). In addition, there was no significant effect of disparity ($\beta = 0.02$, SE = 0.04, $p = 0.66$, $d = 0.01$), nor were there any significant interactions ($\beta = 0.11$, SE = 0.06, $p = 0.10$, $d = 0.06$). Our results are consistent with those of Experiment 1 in that naïve adult performance remains poor for this task. We must conclude that either 2D figural grouping is not responsible for the poor adult performance seen in Experiment 1, or our manipulation was not sufficient to disrupt this grouping effect. Given that the data sets are virtually identical to those obtained in Experiment 1, and the fact that the changes to the stimuli were well above threshold, we believe that the former is the most plausible explanation.

8. Experiment 2e – Scale of configuration

As described in Experiment 1, we used large cartoon characters to capture the interest of young observers. As a result, the stimuli covered a substantial area, spanning a total of 9.7 deg vertically. In that study, although they were asked to maintain fixation on the happy face, observers did have the opportunity to shift attention overtly that is to make eye movements, about the display. It is

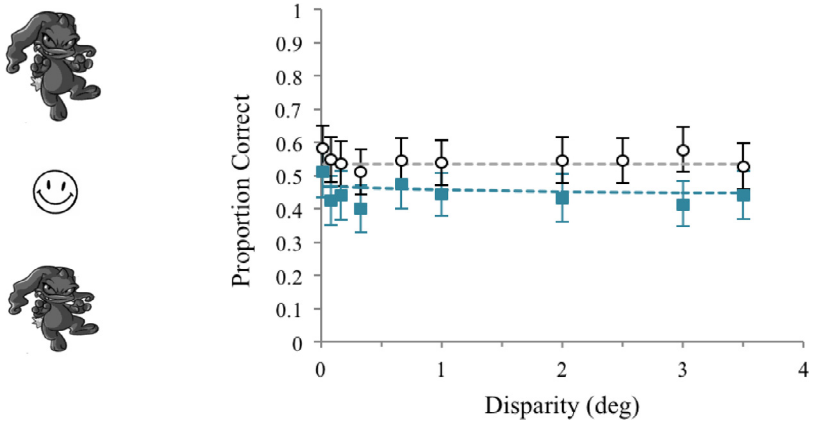

Fig. 5. The variable size stimuli are shown here to illustrate one trial (one eye's image only). The results of the size constancy study are shown to the right (closed squares), along with Experiment 1 adult data (open circles). The dashed lines indicate the average marginal probabilities predicted from the logistic regression model. Error bars indicate one standard error of the mean.

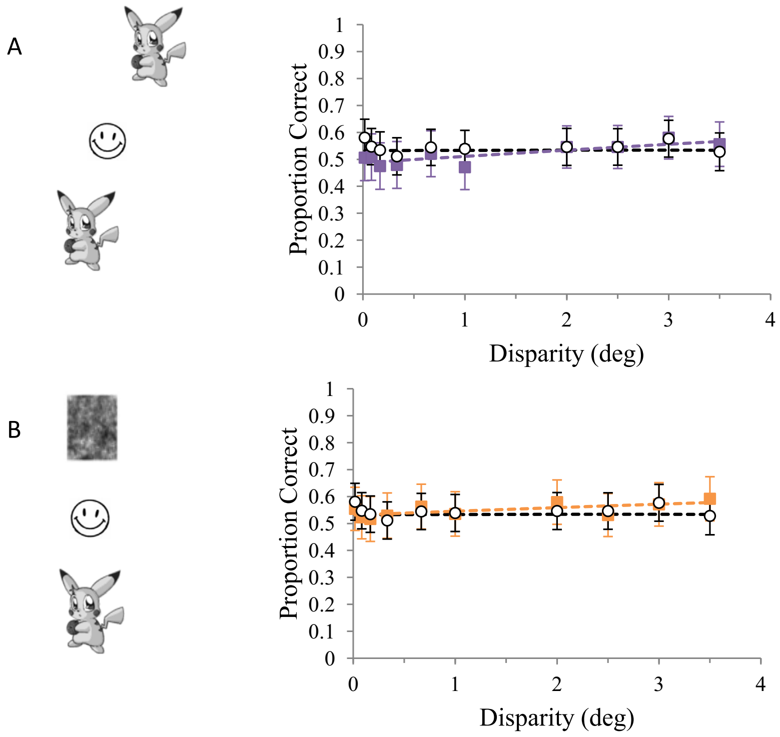

Fig. 6. Stimuli and results of Experiment 2d. A. laterally offset stimuli and B. phase-scrambled stimulus. The graph to the right of each image depicts the results in the lateral offset (closed purple squares) and phase-scrambled (closed orange squares) conditions compared with the adult results from Experiment 1 (open circles). The dashed lines indicate the average marginal probabilities predicted from the logistic regression model and error bars represent one standard error of the mean.

widely known that many aspects of both overt and covert attention develop and are refined with age. A complete summary of the work in the field of the development of attention is outside the scope of this paper, but the interested reader is directed to Mullane, Lawrence, Corkum, Klein, and McLaughlin (2016) for a review, as well as to Enns (1990). Developmental researchers have reported that the ability to focus attention under instruction, with exogenous (cued) stimuli, is mature as young as 6 years of age (Plude, Enns, & Brodeur, 1994). Our study involves exogenous orienting of attention; observers are asked to fixate on the central figure and they know when and where the targets will appear. Therefore, performance differences as a function of age are not likely due to the need for focused attention in general. However, children are slower to re-orient attention (both covertly and overtly), with latencies to initiate eye movements that are almost double those of adults, in children at 8 years (Miller, 1969). Such latencies are likely due to a combination of motor (eye movement) and sensory factors. As a result, younger observers may have been less likely to try to move their attentional 'spotlight' away from the central fixation region than adults. In other words, it is possible that adults were disadvantaged by the large spatial extent of the stimulus because they shifted attention to more peripheral targets.

To evaluate the possible role of the spatial re-allocation of attention in our stimuli, we reduced the scale of the stimulus configuration to a third (keeping relative size and spacing constant). At this scale the entire stimulus display fell within a central 3 deg region and the separation between the fixation marker and closest edges of the targets was reduced to 0.33 deg.

8.1. Methods

Sixteen naïve observers were recruited for this study from the undergraduate participant pool, all of whom had normal or corrected-to-normal visual acuity. They were tested as described in the General Methods section using the scaled version of the stimulus illustrated in Fig. 7. In this configuration, all dimensions of the stimulus were reduced by 0.3, so the edge-to-edge separation was approximately 0.33 deg, as was the width of the characters. The vertical extent of the stimulus configuration was reduced from 9.7 to approximately 3.2 deg. The range of test disparities was the same as used in previous experiments with 8 disparities tested (0.02–3.5 deg) 32 times apiece, in random order, for a total of 256 trials.

8.2. Results and discussion

As is evident from Fig. 7 there was no effect of disparity, nor was there an improvement in accuracy, when viewing the reduced scale versions of our stimuli. These observations were confirmed statistically in the results of the mixed-effect logistic regression that showed no effect of disparity ($\beta = 0.01$, SE = 0.07, $p = 1.00$, $d = 0.01$), or significant difference in accuracy in the reduced scale condition compared to the original full-scale stimulus condition ($\beta = 0.18$, SE = 0.21, $p = 1.00$, $d = 0.10$). There was also no interaction between test conditions as a function of disparity ($\beta = -0.02$, SE = 0.10, $p = 1.00$, $d = 0.01$). As seen in our previous experiments, the scale manipulation introduced here had no impact on observers' performance, and cannot account for the substantial differences in accuracy between adults and children seen in Experiment 1.

Experiments 2a-e show convincingly that the degraded performance observed in Experiment 1 is a reliable phenomenon that is resilient to multiple configural and methodological changes. To summarize, we found that the poor suprathreshold depth discrimination for this two-target configuration seen in our adult group, could not be improved to the child level by stimulus manipulations designed to reduce the effects of the slanted horopter, vergence eye movements, size constancy, figural grouping, or overall stimulus

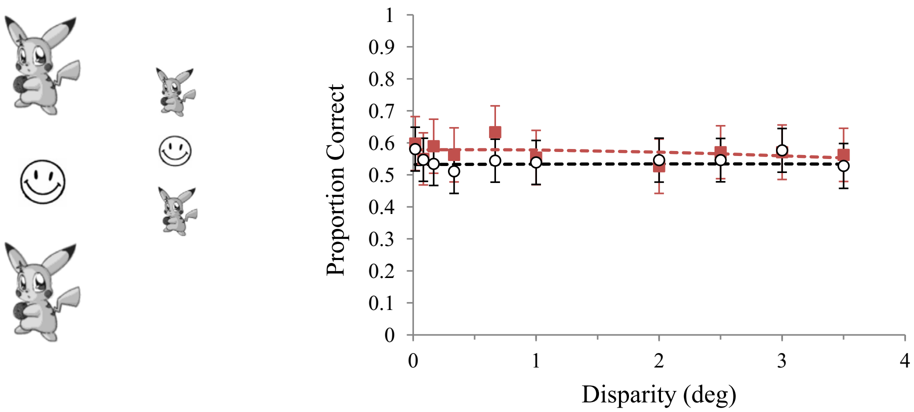

Fig. 7. Illustration of the scaled version of the stimulus used in Experiment 2e (left), and the resultant depth discrimination accuracy (right). Data from the scaled study are plotted as in previous figures along with adult results obtained using the full-scale stimulus from Experiment 1 (open circles). The dashed lines indicate the average marginal probabilities predicted from the logistic regression model. Error bars indicate one standard error of the mean.

scale. Having eliminated these explanations, in Experiment 3 we turn to the possibility that adults are influenced by the 3D spatial layout of the targets.

9. Experiment 3 – 3D spatial layout

In Experiments 1 and 2 observers viewed two stimuli that were simultaneously offset in opposite directions in depth. In Giaschi, Lo et al. (2013) the participants viewed a single target, and judged its depth relative to a fixation plane. The dramatic difference in discrimination accuracy between these studies suggests that the issue may lie with the distribution of the binocular disparity information within the configuration. We evaluate this possibility in two experiments where we a) eliminate one of the stimuli, and therefore present disparity of one sign only and b) evaluate whether the impact of the second target can be modulated voluntarily by presenting the original stimulus and asking observers to judge the depth of one target relative to the fixation marker.

10. Experiment 3a – Single target

10.1. Methods

10.1.1. Participants

Thirteen inexperienced observers were recruited as described in the General Methods section.

10.1.2. Stimuli and apparatus

The apparatus and methods used here were identical to those described in the General Methods section for Experiment 2. Importantly, the stimulus configuration and dimensions were identical to those used in Experiment 1 (including the presence of the happy face fixation marker) except that only one of the figures was presented on each trial (see Fig. 8).

10.1.3. Procedure

The stimulus duration was 750 ms in the training phase and 320 ms in the test phase. On each trial the upper or lower position of the stimulus was selected at random. Observers were asked to indicate whether the single target appeared in front of or behind fixation. As in previous experiments we assessed 10 test disparities, ranging from 0.02 to 3.5 deg with 108 trials per disparity.

10.2. Results and discussion

As shown in Fig. 8 the participants in Experiment 3a performed well above the original adult sample at all disparities, with average performance of 79% correct for the single target compared with 55% correct in the original two target condition. There is no effect of disparity on accuracy, but for the first time we see a substantial improvement in performance compared to our adult observers in Experiments 1 and 2. The mixed-effects logistic regression confirmed that there is a significant increase in accuracy in the single target condition compared to the original two-target condition ($\beta = 1.13$, SE = 0.36, $p < 0.01$, d = 0.62). In addition, there is no significant effect of disparity ($\beta = 0.01$, SE = 0.04, p = 1.00, d = 0.01), nor was there a significant interaction ($\beta = -0.02$, SE = 0.06, p = 1.00, d = 0.01).

It is possible that for the adult observers, the presence of a series of stimuli (upper and lower targets with the central fixation marker), which fall along a linear disparity gradient disrupts their ability to judge the depth sign of one of the targets. However, in addition to manipulating the nature of the disparity information, removal of one of the targets in Experiment 3a also changed the task's attentional demands. While the positional uncertainty could have degraded performance, the presentation of only one target meant that during exposure, the observer had to attend to a smaller region of the display than in the original study. In Experiment 3b, we return to our original two stimulus arrangement, with equal and opposite depth offsets, but modify the task so that observers need only to attend to one area of the display (upper) throughout the study.

11. Experiment 3b – Attentional modulation

If the dramatic improvement in performance seen in Experiment 3a is due to a change in the size of the stimulus area that must be attended to on each trial, then it should be possible to improve performance on the original two target task, simply by asking observers to ignore the lower target and judge the depth of upper target relative to the fixation marker.

11.1. Methods

11.1.1. Participants

Sixteen naïve observers were recruited as described in the General Methods section. Two observers were excluded because it appeared that they reversed the response buttons; their accuracy was consistently near 20%. Results are shown for the remaining 14 observers.

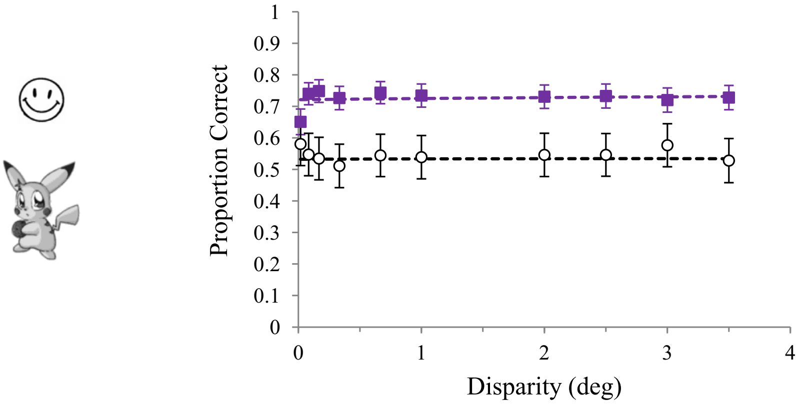

Fig. 8. An illustration of the stimulus configuration used in Experiment 3a (left) and the corresponding depth discrimination results averaged across observers. The single target results (closed squares) are shown along with adult data obtained in Experiment 1 (open circles). The dashed lines indicate the average marginal probabilities predicted from the logistic regression model and error bars represent one standard error of the mean.

11.1.2. Stimuli/apparatus

The stimuli and configuration employed were identical to Experiment 1, except that given the consistent lack of effect of disparity range shown in the preceding experiments we tested a subset of 7 disparities (0.1, 0.17, 0.33, 0.67, 1.0, 2.0, and 3.0 deg) 54 times apiece. Stimuli were displayed for 750 ms during the practice trials, and 320 ms during testing.

11.1.3. Procedure

On each trial two characters were presented (as in Experiment 1), but observers were asked to indicate whether the upper target appeared in front of or behind fixation. They were explicitly instructed to ignore the lower figure.

11.2. Results and discussion

The results of Experiment 3b are shown in Fig. 9 along with results from the original two-target configuration (Experiment 1). As in the original study, observers reported that this task was difficult, and their accuracy was correspondingly poor. A mixed-effects logistic regression confirmed a lack of a difference in accuracy between the attention modulation and the original two-target condition ($\beta = 0.21$, SE = 0.12, $p = 0.21$, $d = 0.12$). In addition, there was no significant effect of disparity ($\beta = 0.03$, SE = 0.04, $p = 0.44$, $d = 0.02$), nor interaction between the effect of the attention modulation as a function of disparity ($\beta = -0.09$, SE = 0.06, $p = 0.21$, $d = 0.05$).

Comparison of our adult observers' performance with that obtained in Experiment 1, shows that adult observers' depth judgements are disrupted even when one of the targets is not attended. As shown in Experiment 3a (Fig. 8) performance improves substantially when one of the targets is removed, but as seen in Fig. 9, observers are unable to improve accuracy by deliberately ignoring one of the figures. This suggests that, for these observers, disparity information is initially processed for the configuration as a whole, which makes segregation of a single feature more difficult.

The results of Experiments 3a and b suggest that it is the distribution of features in depth that disrupts disparity processing for naïve adult observers. It is possible, albeit unlikely, that the issue is simply the simultaneous presence of two targets, irrespective of their displacement in depth. To evaluate this possibility, we conducted a follow-up study ($N = 14$) with the two-target configuration used in Experiment 1, but with both targets displaced in the same direction in depth on each trial. As anticipated, we found there was an improvement in performance (compared to the disparity balanced configuration used in Experiment 1) with an average of 80% correct (Fig. 10). This was confirmed by statistical analysis that revealed a significant increase in accuracy in the 'offset in the same direction' condition compared to the original two-target condition ($\beta = 2.8$, SE = 0.16, $p < 0.001$, $d = 1.54$). Further there was a significant interaction ($\beta = -0.25$, SE = 0.07, $p < 0.001$, $d = 0.14$) however, the small effect size and absence of such effects in our previous experiments makes us hesitant to over interpret this result. On the other hand, the substantial improvement in performance when both targets are offset in the same direction, clearly show that it is not simply the number of targets, but their relative distribution in depth, that influences depth discrimination in our naïve adult observers.

12. General discussion

The results of Experiment 1 confirm that the accurate depth discrimination performance in children as young as 4 years of age reported by Giaschi, Lo et al. (2013) is not due to vergence eye movement strategies, because children also perform very well when the stimuli are equally offset about the fixation plane. Further, this study revealed a striking effect whereby naïve adults consistently performed poorly (often near chance) over this same range of disparities. We investigated the source of this degraded performance in a series of experiments as outlined in Experiments 2 and 3. In Experiment 2, consistent with the work of Harris et al. (2011), we found no significant impact of changing the configuration; nor was there an effect of decreasing exposure duration to 150 ms (thus limiting the potential impact of vergence responses), 2D grouping properties, or overall scale of the stimulus array on adult performance. However, as shown in our third series of experiments, simply removing one of the disparate targets did result in a sizable improvement in performance, an improvement that could not be attributed to attentional factors. This increase in accuracy did not simply reflect the number of targets on the display because displacement of the two targets in the same direction simultaneously also improved adult performance over the full range of disparities. We conclude that the distribution of disparity across targets, was primarily responsible for the poor adult performance in Experiment 1.

12.1. Perceptual organization and disparity processing

When two stimuli were presented simultaneously at equal and opposite disparities, naïve adult observers' discrimination accuracy was poor. This is striking in light of the fact that all observers could reliably distinguish stimuli in the Randot Preschool Stereotest pre-

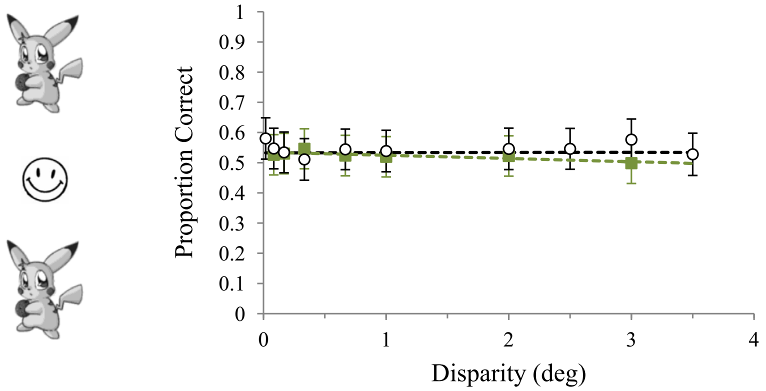

Fig. 9. An illustration of the stimulus configuration used in Experiment 3b (left) and the corresponding depth discrimination results (right). The results (closed squares) are shown along with the adult data obtained in Experiment 1 (open circles). The dashed lines indicate the average marginal probabilities predicted from the logistic regression model and error bars represent one standard error of the mean.

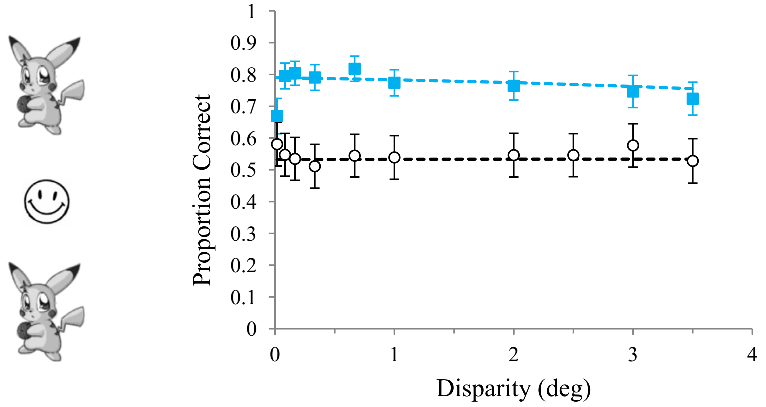

Fig. 10. The depth discrimination results for the follow-up study with both targets displaced in the same direction in depth. The results (closed squares) are shown along with the adult data obtained in Experiment 1 (open circles). The dashed lines indicate the average marginal probabilities predicted from the logistic regression model and error bars represent one standard error of the mean.

sented at 60arcsec (0.02 deg) disparity. Note that in this screening test, a single figure is presented relative to a surround. Our results cannot be fully explained by the 2D properties of the stimuli, their relative size, position, or similarity. While some of these factors may have contributed to the variability in individual performance, it appears that the poor performance of naïve adults is due to the stereoscopic configuration of the stimuli. We hypothesize that this effect is a form of disparity-based grouping that occurs primarily when naïve adult observers view stereoscopic test stimuli. The fact that performance continues to be poor when the task is modified so that observers may ignore the second target (Experiment 3b) suggests that this phenomenon is not under attentional control. Furthermore, this study shows that the improvement seen in Experiment 3a is not due to anticipatory vergence eye movements, since such a strategy would also be applicable here, but no improvement is seen.

While a great deal is known about the basic mechanisms of stereopsis, both physiologically and psychophysically (see Howard, 2012) there are still surprising gaps in our understanding of the nature of depth percepts from stereopsis. Importantly, little is known about how factors such as stimulus configuration influence our interpretation of, and sensitivity to, binocular disparity. One obvious example of this is the fact that stereoscopic depth discrimination improves with the addition of a reference plane. As demonstrated by Westheimer (1979), stereoacuity is degraded when the reference/fixation marker and disparate stimulus are presented sequentially instead of simultaneously. Thus, the temporally coincident reference marker aids relative depth judgements. However, Westheimer and colleagues have also shown that there are numerous spatial properties that interact with disparity processing, including attractive and repulsive effects that are related to element spacing (Westheimer & Levi, 1987). Others have shown configural effects that appear to be due to some form of perceptual grouping (Butler & Westheimer, 1978; Hou, Lu, Zhou, & Liu, 2006; McKee, 1983; Mitchison & Westheimer, 1984; Yin, Kellman, & Shipley, 2000). A similar type of interference has also been shown when the stimulus elements form a biological motion figure (Lu, Tjan, & Liu, 2006). A recent set of experiments by Deas and Wilcox (2014, 2015) have shown that mid-level perceptual grouping of parts to form a closed object reduces the perceived separation of the parts in depth. It is clear from these and other related studies that the percept of depth from disparity is significantly impacted by the configuration of the stimuli, particularly when binocular disparity is isolated from, or in conflict with, monocular cues to depth. The source of this reduction in perceived depth when objects are perceptually grouped is unknown. Mitchison and Westheimer (1984) argue that elevated thresholds reflect inhibition from salient features. McKee and colleagues (McKee, 1983; Vreven, McKee, & Verghese, 2002) have proposed instead that the effects are due to disparity averaging. Deas and Wilcox (2015) propose that object-based disparity smoothing, like that proposed by Marr and Poggio (1976), is the culprit. That is, they argue that the perceptual organization of the stimulus features into a common object guides 'within-object' smoothing operations that promote object cohesion, but make it more difficult to segment individual features. Importantly, they show that continuous gradients of disparity promote such grouping. In a related study, Cammack and Harris (2016) have shown that the stereoscopic system appears to apply smoothing operations based on the gradient of the edges of disparity defined objects. In their study, observers systematically underestimated the amplitude of stereoscopic peaks even though the disparity signal was suprathreshold, and should have been clearly visible. This study, and that of Harris et al. (2011) used random element stimuli which formed (via disparity interpolation) coherent surfaces and objects in depth. By comparison, our isolated cartoon targets and the fixation marker are relatively large, and should appear as three planes or depth steps. It remains somewhat puzzling that we should see such strong grouping effects in this configuration. However, the depth discontinuities in our stimuli may not be salient to our naïve observers; as noted above, further investigation is needed to determine if and how the perceived disparity profile changes with prolonged experience.

We propose that perceptual grouping across disparity is responsible for the poor performance of our adult observers. In all of our experimental manipulations the adult performance was found to be well below that seen in children in Experiment 1 (and in Giaschi, Lo et al., 2013). Only in Experiment 3, when the disparity was not 'balanced' about fixation, did adult performance rise to near that of the children in Experiment 1. This explanation necessitates that children are less susceptible to this disparity-based grouping phenomenon than adults. In their investigations of the development of cue integration, Nardini et al. (2010), examined the effects of sensory integration by combining disparity and texture cues in a slant discrimination task. They found that children performed comparably whether depth cues were presented alone or in combination, however adults had great difficulty judging slant when these cues were presented in combination and in incongruent directions. Nardini et al. (2010) suggest that as the brain matures beyond the age of 8, it gradually develops the ability to integrate different sensory information, which aids our perception when sensory inputs are compatible, but hinders it when they are not. The results of Dekker et al. (2015) provide physiological support for the late maturation of depth cue integration using fMRI techniques. They find that depth from relative motion and stereopsis is not efficiently combined until 11 years of age, and Dekker et al. (2015) provide convincing evidence that this is due to maturation and refinement of neural circuitry.

Similarly, Kovács (2000) proposed that the posterior cortical areas within the occipital lobe, responsible for the spatial integration of objects, develop at a relatively slower rate than other visual pathways. Kovács tested the development of contour integration using Gabor patterns of varying orientation, color, and separation arranged to form contours imbedded in distracter patterns. She found that children between the ages of 5 and 6 showed great difficulty integrating the signals into contours; contour detection improved until well into adolescence. Kovács (2000) also found a developmental difference in the effect of contour spacing on detection. Specifically, the children exhibited greater difficulty in contour detection as spacing increased, while the adults did not. The author suggests that neural circuits responsible for the perception of local features develop prior to the pathways responsible for the integration of these features. In a subsequent study, Kovács (2000) assessed the impact of contextual information on size judgements using the Ebbinghaus illusion. As predicted, she found that children's size estimates were more accurate than those of adults when contextual elements were present. In the context of the configuration used in Experiment 1, it is possible that our young observers were able to interpret the local disparity differences between the fixation and target characters, whereas naïve adults were subject to global grouping that made it difficult to discriminate the depth of the individual characters. Note that the fact that scaling the configuration in Experiment 2e did not improve our adult observers' accuracy means that this phenomenon is configuration-dependent and not simply tied to the spatial dimensions of the array.

12.2. Implications for psychophysical testing

The balanced disparity target arrangement used in Experiment 1 is useful for controlling anticipatory vergence eye movements in studies of stereopsis. However, as we have shown here, it can produce unintended and substantial performance deficits in naïve adult observers. Many, if not most, psychophysical studies of stereopsis in adults involve a small number of highly practiced, experienced observers, so the impact of disparity-based grouping shown here is not often seen or reported. However, the effect of extensive practice on stereoacuity has been documented; for example, Fendick and Westheimer (1983) showed dramatic improvements in stereoacuity in two observers over the first 3000 trials. The improvement in thresholds was greater for targets that were positioned on the peripheral retina (2.5 and 5 deg from fixation). More recently Stransky, Wilcox, and Allison (2014) compared performance on a variety of stereoscopic tasks between groups of observers with different types of training (naïve observers, stereoscopic 3D film professionals, vision scientists with no stereoscopic experience, experienced stereoscopic observers). They found that on most tasks the four groups performed similarly, except for the threshold discrimination task where the experienced stereoscopic observers excelled. Their results echo those of Fendick and Westheimer (1983), but also show that the effects of experience are highly specific to the type of exposure.

As outlined in the Introduction, Harris et al. (2011) also tested a group of naïve observers, using a disparity-balanced configuration. Most relevant to the current study is the fact that, while it was not the focus of their study, only 8 of the 24 observers recruited could perform the disparity-balanced task. As shown here, this is a highly repeatable result. We suggest that the naïve observers in Harris et al. (2011) were also subject to the 3D perceptual grouping effects shown here. In the course of conducting these experiments we found that individuals in the laboratory who were very familiar with psychophysical tests of stereopsis, performed nearly perfectly on the tasks presented here. The impact and nature of training that improves stereoscopic performance is outside the scope of this paper, but is potentially an important issue. It is possible that these disparity-based grouping effects may contribute to anecdotal reports of poor performance in naïve observers on stereoscopic tasks, but often researchers only report data obtained from experienced observers or those who could perform their task, thus excluding the type of effects shown here. This is appropriate if the aim is to illustrate basic properties of the stereoscopic system under ideal conditions; however, it may not reflect how psychophysically naïve observers process disparity information in natural environments. The potential impact of disparity-based grouping should be considered when designing test configurations for use with naïve adult observers.

13. Conclusions

The results of this series of experiments show that stimulus configuration and task demands (simultaneous assessment of equal and opposite disparity offsets), can have a significant impact on perceived depth among adult participants, even at suprathreshold disparities. Consistent with Deas and Wilcox (2014, 2015) we argue that stereoscopic discrimination is affected by disparity-based grouping in naïve adults which makes segmentation difficult. Surprisingly, children do not appear to be subject to the disparity-based grouping phenomenon, this is perhaps due to the relatively slow maturation of cortical circuits which integrate depth information and process global configural information (Dekker et al., 2015; Kovács, 2000; Nardini et al., 2010), which we argue implies a role for visual experience.

Acknowledgments

This work was funded by Natural Sciences and Engineering Council funding to Wilcox and to Giaschi.

We would like to thank Bob Hou for his programming and technical support, Natalie Shaeffer and Alexia Casalinuovo for their help in gathering and organizing the data, and the many students who participated in the experiments.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Birch, E., Williams, C., Drover, J., Fu, V., Cheng, C., Northstone, K., et al. (2008). Randot preschool stereoacuity test: Normative data and validity. *Journal of the American Association for Pediatric Ophthalmology and Strabismus, 12*, 23–26.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis, chapter 7: Converting among effect sizes*. Chichester, West Sussex, UK: Wiley.

Breitmeyer, B., Julesz, B., & Kropfl, W. (1975). Dynamic random-dot stereograms reveal up-down anisotropy and left-right isotropy between cortical hemifields. *Science, 187*(4173), 269–270.

Busettini, C., Fitzgibbon, E. J., & Miles, F. A. (2001). Short-latency disparity vergence in humans. *Journal of Neurophysiology, 95*, 1129–1152.

Butler, T. W., & Westheimer, G. (1978). Interference with stereoscopic acuity: Spatial, temporal, and disparity tuning. *Vision Research, 18*(10), 1387–1392.

Cammack, P. P. K., & Harris, J. M. (2016). Depth perception in disparity-defined objects: Finding the balance between averaging and segregation. *Philosophical Transactions of the Royal Society B, 371*, 20150258.

Ciner, E. B., Schieman, M. M., Schanelklitsch, E., & Weil, L. (1989). Stereopsis testing in 18-month-old to 35-month-old children using operant preferential looking. *Optometry and Vision Science, 66*(11), 782–787.

Cooper, E. A., Burge, J., & Banks, M. S. (2011). The vertical horopter is not adaptable, but it may be adaptive. *Journal of Vision, 11*(3), 1–19.

Cooper, J., Feldman, J., & Medlin, D. (1979). Comparing stereoscopic performance of children using the Titmus, TNO, and Randot stereo tests. *Journal of the American Optometric Association, 50*(7), 821–825.

Coutant, B. E., & Westheimer, G. (1993). Population distribution of stereoscopic ability. *Ophthalmic and Physiological Optics, 13*, 3–7.

Deas, L. M., & Wilcox, L. M. (2014). Gestalt grouping via closure degrades suprathreshold depth percepts. *Journal of Vision, 14*(9), 1–13.

Deas, L. M., & Wilcox, L. M. (2015). Perceptual grouping via binocular disparity: The impact of stereoscopic good continuation. *Journal of Vision, 15*(11), 1–13.

Dekker, T. M., Ban, H., van der Velde, B., Sereno, M. I., Welchman, A. E., & Nardini, M. (2015). Late development of cue integration is linked to sensory fusion in cortex. *Current Biology, 25*, 2856–2861.

Dobson, V., Clifford-Donaldson, C. E., Green, T. K., Miller, J. M., & Harvey, E. M. (2009). Normative monocular visual acuity for early treatment diabetic retinopathy study charts in emmetropic children 5–12 years of age. *Ophthalmology, 116*(7), 1397–1401.

Dowd, J. M., Clifton, R. K., Anderson, D. R., & Eichelman, W. H. (1980). Children perceive large-disparity random-dot stereograms more readily than adults. *Journal of Experimental Child Psychology, 29*, 1–11.

Edwards, M., Pope, D. R., & Schor, C. M. (1999). Orientation tuning of the transient-stereopsis system. *Vision Research, 39*, 2717–2727.

Enns, J. T. (1990). *The development of attention: Research and theory*. New York: Elsevier.

Fendick, M., & Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Research, 23*(2), 145–150.

Foley, J. M., Applebaum, T. H., & Richards, W. A. (1975). Stereopsis with large disparities: Discrimination and depth magnitude. *Vision Research, 15*, 417–421.

Fox, R., Patterson, R., & Francis, E. L. (1986). Stereoacuity in young children. *Investigative Ophthalmology & Visual Science, 27*(4), 598–600.

Giaschi, D., Lo, R., Narasimhan, S., Lyons, C., & Wilcox, L. M. (2013a). Sparing of coarse stereopsis in stereodeficient children with a history of amblyopia. *Journal of Vision, 13*(10).

Giaschi, D., Narasimhan, S., Solski, A., Harrison, E., & Wilcox, L. M. (2013b). On the typical development of stereopsis: fine and coarse processing. *Vision Research, 89*, 65–71.

Harris, J. M., Chopin, A., Zeiner, K., & Hibbard, P. B. (2011). Perception of relative depth interval: Systematic biases in perceived depth. *Quarterly Journal of Experimental Psychology, 65*(1), 73–91.

Helmholtz, H. (1867). *Handbook of physiological optics*. Leipzig: Leopold Voss.

Helmholtz, H. (1962). *Helmholtz's treatise on physiological optics*. New York: Dover Publications.

Heron, G., Dholakia, S., Collins, D. E., & McLaughlan, H. (1985). Stereoscopic threshold in children and adults. *American Journal of Optometry and Physiological Optics, 62*(8), 505–515.

Hess, R. F., & Wilcox, L. M. (1994). Linear and nonlinear filtering in stereopsis. *Vision Research, 34*(18), 2431–2438.

Hess, R. F., & Wilcox, L. M. (2008). The transient nature of 2nd-order stereopsis. *Vision Research, 48*(11), 1327–1334.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65–70.

Hou, F., Lu, H., Zhou, Y., & Liu, Z. (2006). Amodal completion impairs stereoacuity discrimination. *Vision Research, 46*(13), 2061–2068.

Howard, I. P. (2012). *Perceiving in depth*. Oxford: Oxford University Press.

Jones, R. (1980). Fusional vergence: Sustained and transient components. *American Journal of Optometry & Physiological Optics, 57*(9), 640–644.

Koffka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt, Brace, & World.

Kovács, I. (2000). Human development of perceptual organization. *Vision Research, 40*, 1301–1310.

Kovács, I., & Fehér, Á. (1997). Non-Fourier information in bandpass noise patterns. *Vision Research, 37*(9), 1167–1175.

Krekling, S., & Blika, S. (1983). Development of the tilted vertical horopter. *Perception & Psychophysics, 34*(5), 491–493.

Langley, K., Fleet, D. J., & Hibbard, P. B. (1999). Stereopsis from contrast envelopes. *Vision Research, 39*, 2313–2324.

Leat, S. J., St Pierre, J., Hassan-Abadi, S., & Faubert, J. (2001). The Moving Dynamic Random Dot Stereosize test: Development, age norms, and comparison with the Frisby, Randot, and Stereo Smile tests. *Journal of Pediatric Ophthalmology and Strabismus, 38*(5), 284–294.

Lu, H., Tjan, B. S., & Liu, Z. (2006). Shape recognition alters sensitivity in stereoscopic depth discrimination. *Journal of Vision, 6*, 75–86.

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science, 194*(4262), 283–287.

McKee, S. P. (1983). The spatial requirements for fine stereoacuity. *Vision Research, 23*(2), 191–198.

McKee, S. P., Verghese, P., & Farell, B. (2004). What is the depth of a sinusoidal grating? *Journal of Vision, 4*(7), 524–538.

McKee, S. P., Verghese, P., & Farell, B. (2005). Stereo sensitivity depends on stereo matching. *Journal of Vision, 5*(10), 783–792.

Meier, K., Giaschi, D., Wilcox, L. M., Seemiller, E., & Candy, T. R. (2016). Vergence responses to fine and coarse disparities: Adult-like tuning functions at 5 years of age. *Journal of Vision, 16*(12), 841–841.

Miller, L. K. (1969). Eye-movement latency as a function of age, stimulus uncertainty, and position in the visual field. *Perceptual and Motor Skills, 28*, 631–636.

Mitchell, D. E. (1969). Qualitative depth localization with diplopic images of dissimilar shape. *Vision Research, 9*, 991–994.

Mitchison, G. J., & Westheimer, G. (1984). The perception of depth in simple figures. *Vision Research, 24*(9), 1063–1073.

Mullane, J. C., Lawrence, M. A., Corkum, P. V., Klein, R. M., & McLaughlin, E. N. (2016). The development of and interaction among alerting, orienting, and executive attention in children. *Child Neuropsychology, 22*(2), 155–176.

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences, 107*(39), 17041–17046.

Ogle, K. N. (1952). Disparity limits of stereopsis. *Archives of Ophthalmology, 48*(1), 50–60.

Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America, 43*, 906–913.

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica, 86*, 227–272.

Rashbass, C., & Westheimer, G. (1961). Disjunctive eye movements. *Journal of Physiology, 159*, 339–360.

Regan, D. (1988). Low contrast letter charts and sinewave grating tests in ophthalmological and neurological disorders. *Clinical Vision Science, 2*, 235–250.

Romano, P. E., Romano, J. A., & Puklin, J. E. (1975). Stereoacuity development in children with normal binocular single vision. *American Journal of Ophthalmology, 79*(6), 966–971.

Schor, C. M., Edwards, M., & Sato, M. (2001). Envelope size tuning for stereo-depth perception of small and large disparities. *Vision Research, 41*(20), 2555–2567.

Simons, K. (1981). Stereoacuity norms in young children. *Archives of Ophthalmology, 99*(3), 439–445.

Smallman, H. S., & MacLeod, D. I. A. (1994). Size–disparity correlation in stereopsis at contrast threshold. *Journal of the Optical Society of America A, 11*(8), 2169–2183.

Stransky, D., Wilcox, L. M., & Allison, R. S. (2014). Effects of long-term exposure on sensitivity and comfort with stereoscopic displays. *ACM Transactions on Applied Perception, 11*(1), 2–13.

Tomac, S., & Altay, Y. (2000). Near stereoacuity: Development in preschool children; normative values and screening for binocular vision abnormalities; a study of 115 children. *Binocular Vision & Strabismus Quarterly, 15*, 299–321.

Tschermak, A., & Hoefer, P. (1903). Ueber binoculare Tiefenwahrnehmung auf Grund von Doppelbildern. *European Journal of Physiology, 98*, 299–321 (in German).

Tyler, C. W. (1973). Stereoscopic vision: Cortical limitations and a disparity scaling effect. *Science, 181*, 276–278.

Tyler, C. W., & Scott, A. B. (1979). Binocular vision. *Physiology of the Human Eye and Visual System*, 643–671.

Vreven, D., McKee, S. P., & Verghese, P. (2002). Contour completion through depth interferes with stereoacuity. *Vision Research, 42*, 2153–2162.

Westheimer, G. W. (1979). Cooperative neural processes involved in stereoscopic acuity. *Experimental Brain Research, 35*, 585–597.

Westheimer, G., & Levi, D. M. (1987). Depth attraction and repulsion of disparate foveal stimuli. *Vision Research, 27*(8), 1361–1368.

Westheimer, G., & Mitchell, D. E. (1969). The sensory stimulus for disjunctive eye movements. *Vision Research, 9*, 749–755.

Wheatstone, C. (1938). Contributions to the physiology of vision-part the first. On some remarkable and hitherto unobserved phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London, 128*, 317–394.

Wilcox, L. M., & Allison, R. S. (2009). Coarse-fine dichotomies in human stereopsis. *Vision Research, 49*(22), 2653–2665.

Wilcox, L. M., & Hess, R. F. (1995). Dmax for stereopsis depends on size *not* spatial frequency content. *Vision Research, 35*, 1061–1069.

Wilcox, L. M., & Hess, R. F. (1996). Is the site of non-linear filtering in stereopsis before or after binocular combination? *Vision Research, 36*, 391–399.

Wilcox, L. M., & Hess, R. F. (1997). Scale selection for second-order (non-linear) stereopsis. *Vision Research, 37*, 2981–2982.

Yang, Q., Bucci, M. P., & Kapoula, Z. (2002). The latency of saccades, vergence, and combined eye movements in children and in adults. *Investigative Ophthalmology and Vision Science, 43*(9), 2939–2949.

Yin, C., Kellman, P. J., & Shipley, T. F. (2000). Surface integration influences depth discrimination. *Vision Research, 40*, 1969–1978.

Zeigler, H. P., & Leibowitz, H. (1957). Apparent visual size as a function of distance for children and adults. *The American Journal of Psychology, 70*(1), 106–109.

Ziegler, L. R., & Hess, R. F. (1997). Depth perception during diplopia is direct. *Perception, 26*, 1225–1230.