Overestimation of the number of elements in a three-dimensional stimulus

Saori Aida — Graduate School of Marine Science and Technology, Tokyo University of Marine Science and Technology, Koto-ku, Tokyo, Japan ✉

Tsutomu Kusano — Graduate School of Marine Science and Technology, Tokyo University of Marine Science and Technology, Koto-ku, Tokyo, Japan

Koichi Shimono — Graduate School of Marine Science and Technology, Tokyo University of Marine Science and Technology, Koto-ku, Tokyo, Japan

Wa James Tam — School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada

Observers' numerosity judgments in binocular stereopsis were examined in four experiments, using random-dot stereograms (RDSs) that depicted a two-dimensional (2-D) stimulus side-by-side with a three-dimensional (3-D) stimulus. When the RDSs were correctly fused, a single surface and two (or three) transparent surfaces were observed for the 2-D and 3-D stimuli, respectively. Observers completed a numerosity discrimination task, where they judged which of the two stimuli had a greater number of dot elements. Results showed that (a) the 3-D stimulus was judged to contain more elements than the 2-D stimulus, even when both had the same number of elements, (b) the amount of overestimation increased as a function of the number of elements and the binocular disparity between the front and back surfaces of the 3-D stimulus, (c) the ratio of the physical number of elements in the front surface to that in the back surface of the 3-D stimulus had no effect on the magnitude of overestimation, and (d) when the number of elements for the two surfaces were judged separately, the ratio had more effect on the judged number of elements in the back surface than in the front surface. These results indicate that the extent of overestimation in the numerosity judgment of a set of elements in a stimulus depends on the number of depth layers in which the elements are embedded.

Introduction

We report here a phenomenon in which a random-dot stereoscopic (RDS) three-dimensional (3-D) stimulus is perceived to contain more dot elements than a stereoscopic two-dimensional (2-D) stimulus, when both stimuli have the same number of elements. The fact that the same physical number of elements is judged to be different between the 3-D and 2-D stimuli suggests that numerosity is not perceived independently from the depth structure of the stimuli. Thus, examining conditions that can give rise to an overestimation of the number of elements by comparing 2-D and 3-D stimuli will help increase our understanding of the underlying mechanism(s) for numerosity judgment.[1]

Researchers have paid much attention to the study of numerosity judgment using 2-D stimuli with elements that are distributed in a flat surface, and have neglected the use of 3-D stimuli where elements are perceived at different depths.[2] It is well documented that humans can estimate fairly accurately the number of elements presented in a flat surface, even if the number is relatively large (e.g., Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Krueger, 1984; Tibber, Greenwood, & Dakin, 2012). For example, the number of elements presented in a flat surface is described by Krueger (1984) as a power function with

Citation: Aida, S., Kusano, T., Shimono, K., & Tam, W. J. (2015). Overestimation of the number of elements in a three-dimensional stimulus. *Journal of Vision*, *15*(9):23, 1–16, doi:10.1167/15.9.23.

doi: 10.1167/15.9.23 Received February 6, 2015; published July 31, 2015 ISSN 1534-7362 © 2015 ARVO

exponents less than unity for an absolute judgment task. Furthermore, the accuracy and precision of numerosity judgment for a discrimination task are known to be affected by various stimulus properties, such as element density, element size, luminance of elements, size of the stimulus containing the elements, and location of retinal stimulation (e.g., Burr & Ross, 2008; Dakin et al., 2011; Ginsburg & Nicholls, 1988; Ross & Burr, 2010; Valsecchi, Toscani, & Gegenfurtner, 2013).

In contrast, only a few studies have examined the properties of numerosity judgment for a stimulus with depth (Aida, Kusano, & Shimono, 2013; Bell, Manson, Edwards, & Meso, 2015; Schütz, 2012). This suggests that there might be an underlying assumption among researchers that numerosity judgment of a 2-D stimulus is essentially the same as that for a 3-D stimulus (see Bell et al., 2015). Recently, however, Schütz (2012) reported that the number of dot elements in the back surface of a 3-D stimulus consisting of two motion-transparent planes was overestimated compared to that in the front surface. His stimulus consisted of two overlaid surfaces, either with relative disparity or with zero disparity, created with dots that moved in two distinct directions for the two surfaces. Aida et al. (2013) reported that the total number of elements in a stereoscopic 3-D stimulus, which depicted either two or three overlaid surfaces, was overestimated in comparison with that in a stereoscopic 2-D stimulus, which depicted a single flat surface. The reports of an overestimation of the back-surface element (Schütz, 2012; back-surface-element overestimation phenomenon) and of the total number of elements (Aida et al., 2013; total-element overestimation phenomenon) suggest that depth can affect numerosity judgment.

To explain back-surface-element overestimation, Schütz (2012) used Tsirlin, Allison, and Wilcox's (2012) hypothesis, which was proposed to explain the phenomenon that the back surface of a stereoscopic two-surface stimulus is perceived as being denser than its front surface when both surfaces have the same number of dot elements. In their model, a higher order process assigns the dots in the back surface to their surrounding blank areas to form an opaque background surface. The hypothesis assumes that the increase in neural activity created by this process results in the perception of a denser back surface. Although Tsirlin et al. (2012) asked observers to compare the dot density and not the numerosity, their back-surface bias hypothesis is, in general, consistent with the idea that the numerosity overestimation of a 3-D stimulus is caused by an overestimation of elements in its back surface. If the total-element overestimation phenomenon of a 3-D stimulus is caused by an overestimation of elements in its back surface because of a perceived denser surface, then the phenomenon can be explained by the back-surface bias hypothesis.

We conducted four experiments using overlaid surfaces to examine the total-element overestimation phenomenon (Aida et al., 2013) and to determine the role that potential factors might play in the phenomenon. Observers were asked to perform a discrimination task by comparing the number of elements on either a 2-D or a 3-D stimulus with that of a 3-D stimulus; the stimuli were displayed side-by-side. In Experiment 1, the number of elements was manipulated to examine its impact on the overestimation. In Experiment 2, binocular disparity was manipulated. We expected that if the total-element overestimation phenomenon is affected by the amount of depth in a 3-D stimulus, then binocular disparity would have an impact on perceived numerosity. In Experiments 3 and 4, we manipulated the ratio of elements on the front surface to that of the back surface of 3-D stimuli with overlaid surfaces while keeping the total number of elements constant to examine the hypothesis that the back surface plays a role in the total-element overestimation phenomenon.

General methods

Apparatus

A computer (Dell Dimension 9100) generated the test stimuli that were displayed on a 23-in. monitor (Diamond Crusta RDT23IWLM-S, Mitsubishi) with a resolution of 1024 × 768 pixels. The monitor was set such that the center of the display was at the eye level of the seated observer. The observer's head was supported by a head-and-chin rest and he or she viewed the stimuli with anaglyph glasses from a distance of 60 cm. The experimental room was completely dark except for the light from the monitor that provided dim illumination.

Stimuli

As illustrated in Figure 1A, the stimuli were RDSs that consisted of rectangular elements. The size of each element was 6.8 × 12.0 arc min. The size of each RDS was approximately 11.5 × 15.5 arc deg in Experiments 1, 3, and 4. It was smaller (6.6 × 7.2 arc deg) in Experiment 2 in which eye movements were controlled with a fixation cross. The luminance of each element of the RDS was 0.3cd/m^2 and the background was 37.0 cd/m^2. The luminance of the stimuli was measured with a luminance meter (LS100, Konica Minolta, Inc., Tokyo, Japan).

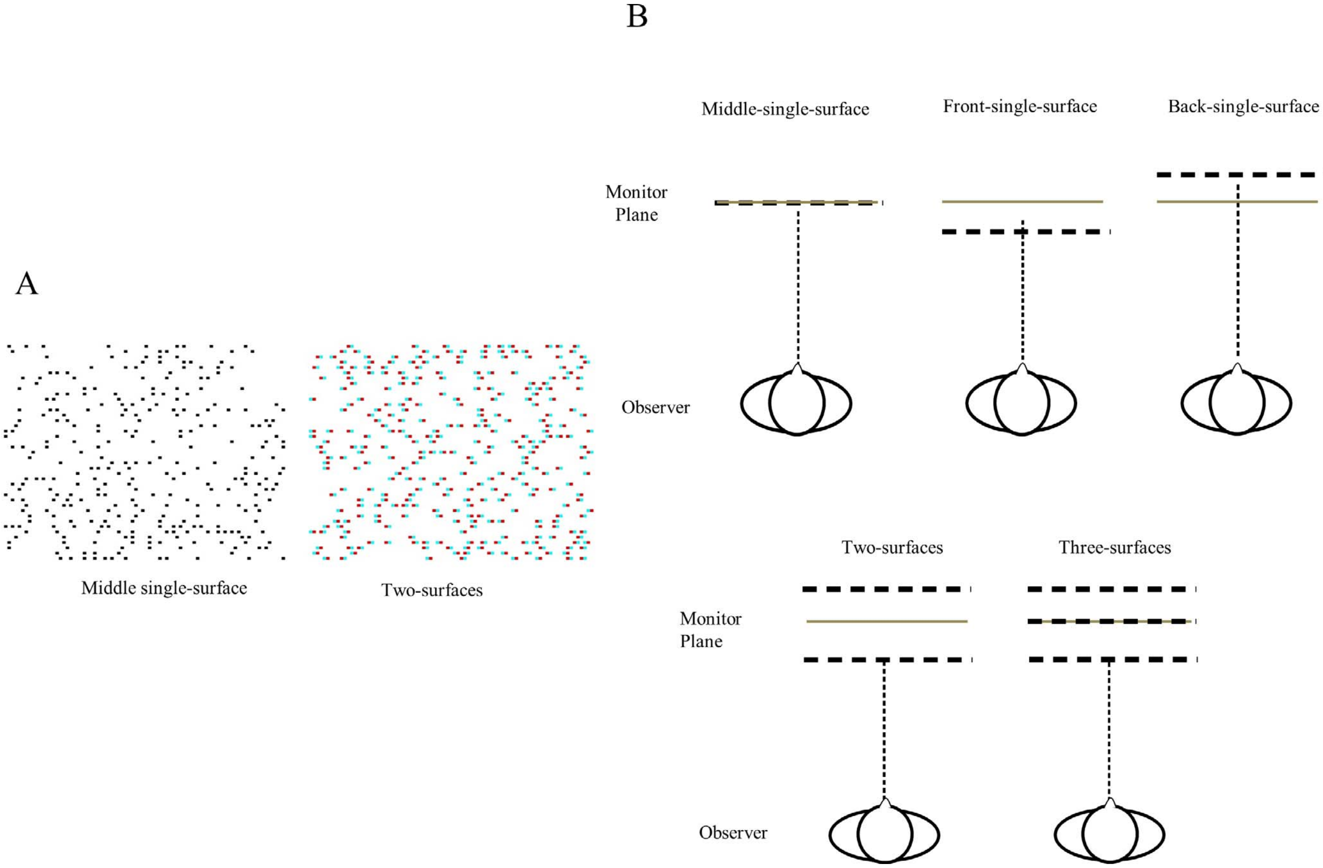

Figure 1. (A) Illustration of typical 3-D stimuli used in the present study. Readers might perceive the total-element overestimation phenomenon using anaglyph glasses and comparing the number of elements for the two-surface stimulus against the middle-single-surface stimuli. (B) Schematic illustration of observers' perception of the stimuli when properly fused. Upper three graphics illustrate the top view for three types of single-surface stimuli: (starting from the left) for the middle-, front-, and back-single-surface stimuli, with representations of observers' perception of a single surface at the monitor plane, at a plane in front of the monitor, and at a plane behind the monitor, respectively. Lower two graphics illustrate the top view of a two-surface stimulus and a three-surface stimulus. Observers would perceive two and three stereo-surfaces in the same visual direction, respectively.

There were two sets of RDS stimuli: a 2-D stimulus and a 3-D stimulus. The 2-D stimulus, when fused, depicted a single surface of elements that would be perceived at the monitor plane (Experiments 1 through 4) and also at a front-parallel plane with crossed or uncrossed disparity with respect to the monitor plane (Experiments 3 and 4; see upper three graphics in Figure 1B). We will refer to the former as a middle-single-surface stimulus and to the latter as either a front-single-surface or a back-single-surface stimulus, respectively. The 3-D stimulus, when fused, depicted two or three overlaid surfaces of elements (see lower two graphics in Figure 1B) that would be perceived as multiple surfaces at different depths in the same visual direction (i.e., stereo transparency; e.g., Aida, Shimono, & Tam, 2015; Akerstrom & Todd, 1988; Julesz, 1971; Tsirlin, Allison, & Wilcox, 2008, 2012). We will refer to a stimulus with two overlaid surfaces as a two-surface stimulus and that with three overlaid surfaces as a three-surface stimulus. The two-surface and three-surface stimuli were used in Experiment 1 and a two-surface stimulus was used in Experiments 2, 3, and 4. The positions of the elements in the surfaces of all the stimuli were randomly assigned and manipulated so that adjacent elements within a single depicted surface and between two depicted surfaces did not overlap or contact each other. The position of the elements in the multi-surfaces stimulus was also manipulated so that the disparity gradient was less than unity in order to ensure binocular fusion (see Burt & Julesz, 1980; Howard & Rogers, 2012). For that purpose, we excluded any two dots appearing within 8.0 arc min in width and 6.8 arc min horizontal separation. Stimuli were displayed laterally on both sides of the midsagittal plane that was aligned with the midline of the monitor, and the centers of the stimuli were horizontally separated by 17.0 arc deg in Experiments 1, 3, and 4, and by 8.8 arc deg in Experiment 2. In the fixation condition of Experiment 2, a cross was placed at the center between the two stimuli. The total disparities (interplane disparities summed up) of the 3-D stimuli used in the present study were well within the fusional range of observers for the RDS (Howard & Rogers, 2012; Yeh & Silverstein, 1990) and within the range in which stereo transparency can be observed (e.g., Tsirlin et al., 2008). In the without-fixation condition of Experiment 2, the stimuli were the same as those used in the fixation condition minus the fixation cross.

Procedure

For each trial in each experiment, observers were asked to indicate which of the two stimuli, presented side by side on the monitor, had a larger number of elements. The stimuli were presented without a time limit and remained visible until the observers finished responding. Except for the with-fixation condition in Experiment 2, observers were allowed to move their eyes. Experiments 1, 3, and 4 had a within-subject design and Experiment 2 had a between-subjects design. Observers were allowed to take a break after each block of trials if their eyes were tired.

Observers were screened for stereopsis before each experiment. They were asked to verbally report the magnitude of perceived depth in millimeters of a two-surface RDS for three different binocular disparities (4.0, 8.0, and 12.0 arc min). Each stimulus was presented once and with a different presentation order for each observer. We plotted the reported depth as a function of disparity and calculated the slopes of the regression lines for the plotted data for each observer. Observers were allowed to participate only if their slope was larger than zero. All four experiments consisted of practice and experimental sessions. In the practice session, observers performed several training trials, which were randomly selected from those used in the experimental session, till the experimenter judged that observers understood the task.

Observers

Thirty-two students from the university community, ranging in age from 19 to 26 years, participated in the study; two were excluded from Experiment 2 after failing the test for stereopsis. Experiment 1 was conducted with one author (female) and six naive observers (two males), Experiment 2 with 16 naive observers (14 males), Experiment 3 with one author (female) and seven naive observers (five males), and Experiment 4 with one author (female) and six naïve observers (two males). Five observers participated in more than two experiments; one participated in Experiments 1 through 4, two in Experiments 1, 3, 4, and one in Experiments 1, 2, and 4, and one in Experiments 1 and 4. All had normal or corrected-to-normal visual acuity. Observers gave informed consent prior to taking part in the experiments, which were conducted in accordance with the ethical principles embedded in the Declaration of Helsinki.

Psychophysical data analysis

From the data obtained from Experiments 1 through 4, we calculated the point of subjective equality (PSE) for each observer and each condition. PSE was defined as the number of elements on the comparison stimuli that produced the same perceived number as that of the standard stimuli. It was calculated from a psychometric function fitted to the percentage of the responses in which the number of elements on the comparison was perceived to be larger than that of the standard, as a function of number of elements on the comparison. The psychometric function was a logistic function that was fitted using Sigmaplot 11.2 (Systat Software, Inc., San Jose, CA), and the number of elements on the comparison that produced 50% response in the function was identified as the PSE. Furthermore, we determined a bias for the PSE by subtracting the PSE value from the number of elements on the standard; when the bias was positive, overestimation in numerical judgment had occurred, and when negative, underestimation had occurred.

Experiment 1

In Experiment 1, we manipulated the number of elements in the 2-D stimulus to confirm the overestimation phenomenon (Aida et al., 2013) and to estimate the number of elements in the 3-D stimulus that is perceived to be the same as that of the 2-D stimulus. We also compared the perceived numerosity of a two-surface stimulus and that of a three-surface stimulus to examine whether the overestimation can be observed between two 3-D stimuli.

Method

Stimuli

When the number of elements in a 2-D stimulus was compared against those in a 3-D stimulus, a middle-single-surface stimulus was used as a standard and a two- or three-surface stimulus was used for comparison. As well, when the number of elements was compared with respect to two 3-D stimuli, a two-surface stimulus was used as a standard and a three-surface stimulus was used for comparison. The number of elements on the standard stimulus consisted of 72, 150, 300, or 600, while the total number of elements on the comparison stimuli was varied, in incremental step sizes of 18, 30, 42, and 60 elements, from 36–108, 90–210, 216–384, and 480–720, respectively. Thus, for each level of the standard, there were five numbers of elements used for the comparison (e.g., for a standard with 150 elements, the five stimuli used for comparison had 90, 120, 150, 180, and 210 elements). For the two-surface stimulus, each surface consisted of one half of the total number of elements; for the three-surface stimulus, each consisted of one third the total number of elements. The total disparity (interplane disparities summed up) of the 3-D stimulus or the comparison was 12.0 arc min; the disparity pairs used to generate the stimuli were 6.0 and −6.0 arc min with respect to the monitor plane for a two-surface stimulus and 6.0, 0, and −6.0 arc min for a three-surface stimulus. Positive and negative values represented crossed and uncrossed disparities, respectively.

Procedure

The experiment consisted of three sessions. In each session, there were three surface combinations of the standard and the comparison: (a) a middle-single-surface (monitor plane) versus a two-surface combination, (b) a middle-single-surface versus a three-surface combination, and (c) a two-surface versus a three-surface combination. In each session there were four blocks, in which the standard stimuli consisted of 72, 150, 300, and 600 elements; the presentation order of the four blocks differed among observers. In each block, the number of elements on the comparison and its presentation location were randomly selected from five different numbers of elements and two locations (right or left), respectively, with five repetitions. Thus, there were 600 trials (3 surface combinations × 4 numbers of elements on the standard × 5 numbers of elements on the comparison × 2 locations × 5 repetitions) in total for each observer.

Results

We conducted a two-way repeated measures ANOVA (4 numbers of elements on the standard × 3 surface combinations) on the bias of the PSE. The analysis showed that the main effects of the number of elements, $F(3, 18) = 17.44$, $p < 0.01$, and surface combinations, $F(2, 12) = 4.05$, $p < 0.05$, were statistically significant, but their interaction was not. The significant main effects of the number of elements can be seen in Figure 2, which shows the mean biases of the PSE as a function of the number of elements on the standard with the surface combinations as the parameter. As can be seen in the figure, the bias increased as a function of the number of elements for every surface combination. Post hoc analyses (Tukey tests) showed that the difference of the mean bias between each number of elements was statistically significant ($p < 0.05$), except for those between 72 and 150 and between 300 and 600 elements. Post hoc analyses (Tukey tests), however, showed that there were no statistical differences between any pairs of the three surface combinations. To examine the effect of the surface combination on the bias of the PSE further, we averaged the biases among the four numbers of elements and performed a one-way repeated measures ANOVA (three surface combinations) on the average biases and found no significant difference in them. These statistical results can also be seen in Figure 2, suggesting that the bias of the PSE increased similarly across the range of the number of elements on the standard for the three surface combinations.

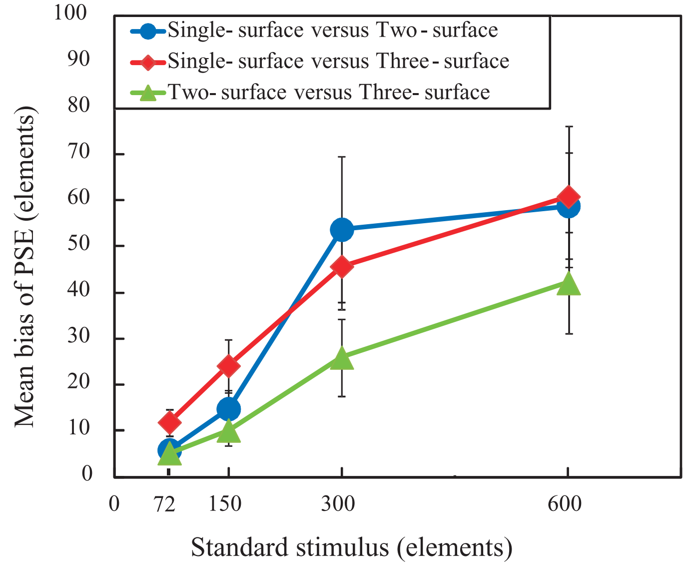

Figure 2. Results from Experiment 1. Mean biases of the PSE as a function of the number of elements in the standard stimulus. Blue circles, red diamonds, and green triangles indicate the extent of the bias for the single-surface versus two-surface combination, for the single-surface versus three-surface combination, and for the two-surface versus three-surface combination, respectively. Error bars are ±1 *SE*.

Figure 2 also shows that the sign of the mean bias of the PSE is positive for every surface combination and every number of elements, suggesting that the overestimation phenomenon has occurred in this experiment.[3] This suggestion is consistent with the result of the two-way repeated measures ANOVA, discussed above, showing that the intercept was significantly different from zero, $F(1, 6) = 24.69$, $p < 0.01$. This significance and the positive sign of the smallest mean bias suggest that the stereoscopic 3-D stimulus was perceived to contain more elements than the stereoscopic 2-D stimulus. Furthermore, the result indicates that the overestimation phenomenon also occurs for the two-surface versus three-surface combination, showing that a 3-D stimulus with three surfaces was judged to contain more elements than that with two surfaces when both had the same number of elements and the disparity between their two outermost surfaces was the same.

Discussion

The present study confirmed the findings of Aida et al. (2013) who reported the total-element overestimation phenomenon for 3-D numerosity judgments. As can be seen in Figure 2, for a relatively large range of the physical number of elements (72–600) that was manipulated, the perceived number of elements in a 3-D stimulus with multiple depth layers was overestimated compared to that in a 2-D standard stimulus containing the same number of elements. The extent of overestimation is approximately between 7% and 16% of the number of elements embedded in the standard stimulus. Furthermore, the overestimation also occurred when comparing two 3-D stimuli with multiple depth layers, specifically when comparing a two-surface with a three-surface 3-D stimulus.

The total-element overestimation phenomenon reported in the present study as well as the back-surface-element overestimation phenomenon reported in Schütz (2012), however, are apparently at odds with the results reported by Bell et al. (2015), who found that the number of elements distributed in depth across a cylindrical volume was perceived to be the same as that distributed in a flat surface. At the present, we think that the differences may be due to a difference of the properties in the 3-D stimuli used. In this study and Schütz's study, the 3-D stimulus was a stereoscopic and/or motion transparent stimulus, which produced two or three overlaid surfaces, while in Bell et al., the 3-D stimulus was one that produced a solid volumetric impression. If the difference in the results is caused by the different stimuli used, it suggests that surfaces overlapping in depth play a key role on the total-element- and back-surface-element overestimation phenomena (see similar discussion in Bell et al., 2015).

Because the total-element overestimation phenomenon is observed when a stereoscopic 3-D stimulus with multiple layers and a 2-D stimulus are compared, one might wonder whether stereoscopic size constancy could have played a role in the findings. In stereopsis, an object that is perceived to be behind another is known to appear to be larger than that perceived to be in front when their retinal images are the same size, (e.g., Howard & Rogers, 2012; Oyama, 1974). Given that in the current study the physical sizes of the stimulus area and elements for the front surface are the same as those of the back surface, the perceived sizes of the stimulus area and elements in the back surface could have appeared to be larger than those in the front surface. Furthermore, it has been suggested that the size of a stimulus area where the elements are presented and the size of each element can affect numerosity judgments (e.g., Dakin et al., 2011; Tibber et al., 2012; Tokita & Ishiguchi, 2010; cf. Allik, Tuulmets, & Vos, 1991). As a matter of fact, when either the size of a stimulus area or each element is increased, the number of elements is likely to be overestimated (e.g., Dakin et al., 2011). However, according to geometrical calculations, the expected size difference between the elements perceived in the front and back surfaces, which can produce the same size of retinal images of the elements used in this experiment, were 0.4 mm (2.2 arc min) in height and 0.6 mm (3.4 arc min) in width, and as such are rather small. In fact, seven observers reported difficulty in seeing a clear size difference between the perceived front and back elements. Based on the geometrical analysis and the observers' verbal report, we believe that a change in the perceived size of the surfaces or elements, even if it did occur, most likely had a negligible effect on the perceived numerosity in this experiment. In Experiment 3 we will present evidence that shows why stereoscopic size constancy and a perceived increase in size of the elements cannot account for the overestimation phenomenon.

Moreover, one might argue that the total-element overestimation phenomenon is a byproduct of vergence eye movements. For example, if the eyes overconverge or overdiverge so that elements in a front or back surface of the 3-D stimulus could not fuse, then double vision would occur for the 3-D stimulus, and its perceived number of elements would increase. However, the possibility of double vision is low in this experiment because the only visible and high-contrast stimuli were the 2-D and 3-D stimuli themselves and the eyes were likely to be converged on the stimuli, and because the disparity values of the stimuli used were well within the fusional area (Howard & Rogers, 2012; Yeh & Silverstein, 1990). Nevertheless, because eye positions were not controlled in this experiment, we added a visual fixation stimulus to minimize vergence eye movements in Experiment 2.

Experiment 2

In Experiment 2, we examined the effect of binocular disparity on the total-element overestimation phenomenon. We hypothesized that manipulating binocular disparity can have an effect on the phenomenon because the phenomenon was observed with a (stereoscopic) 3-D stimulus. We also examined the effect of the presence or absence of a fixation stimulus on the phenomenon. Given that the binocular disparities used were well within the binocular fusional area, the presence or absence of a fixation stimulus should not have an effect on the phenomenon, and this finding would support our argument that vergence eye movements and double vision did not contribute to the results that were obtained in Experiment 1.

Method

Stimulus

In this experiment, a middle-single-surface (2-D) stimulus was used as a standard and a two-surface (3-D) stimulus was used as a comparison. The number of elements on the 2-D standard was either 150 or 300. The number of elements on the 3-D comparison was varied between 130 and 170, with an incremental step size of 10 elements (against the 150-element standard) and between 260 and 340, with an incremental step size of 20 elements (against the 300-element standard). Thus, for each level of the standard, there were five numbers of elements used for the comparison. For example, for the standard with 150 elements, the five stimuli used for comparison had 130, 140, 150, 160, and 170 elements. For the two-surface comparison, each surface consisted of one half of its total number of elements. The total disparities (interplane disparities summed up) of the 3-D comparison stimulus were 4.0, 8.0, and 12.0 arc min; the disparity pairs used to generate the stimuli were 4.0 and 0.0 arc min, 4.0 and –4.0 arc min, and 8.0 and –4.0 arc min. Positive and negative values represented crossed and uncrossed disparities, respectively, with respect to the monitor plane. Fixation was controlled with a fixation cross that was placed at the center between the standard and the comparison in one (with fixation) condition.

Procedure

There were two experimental sessions, one using a standard with 150 elements and the other with 300 elements. Eight observers completed each session. There were two blocks of trials per session: a with-fixation condition and without-fixation condition. The order of the fixation conditions was counterbalanced. In the with-fixation condition, observers were asked to fixate the fixation cross presented on the monitor (see General methods) and in the without-fixation condition they were allowed to move their eyes. In each block, there were three binocular disparity conditions (4.0, 8.0, and 12.0 arc min) and the order of the disparity conditions was varied among observers. In each disparity condition, the number of elements on the comparison and its presentation locations were randomly selected from five different numbers of elements and two locations (right or left), respectively, with five repetitions. Thus, in total there were 300 trials (2 fixation conditions $\times$ 3 disparities $\times$ 5 numbers of elements on the comparison $\times$ 2 locations $\times$ 5 repetitions) for each observer.

Results

We conducted a three-way mixed-design ANOVA (2 number of elements on the standard $\times$ 2 fixation conditions $\times$ 3 binocular disparities) on the bias of the PSE, with the number of elements as between-subjects variable and with disparity and fixation as within-subject variables. The main effect of binocular disparity, $F(2, 28) = 4.08$, $p < 0.05$, was statistically significant, whereas the main effects of fixation and number of elements on the standard stimuli, and all their interactions were not. The results can be seen in Figure 3, which shows the mean bias of the PSE as a function of the binocular disparity of the standard, with the number of elements as the parameter for the with-fixation condition and for the without-fixation condition; the data were plotted separately for the conditions with 150 elements (Figure 3A) and 300 elements (Figure 3B). As can be seen in Figure 3, the bias increased as a function of binocular disparity except at the largest disparity for the condition with 150 elements; the bias was relatively constant between the two fixation conditions. Post hoc analyses (Tukey tests) showed that the mean bias is statistically significant between 4.0 and 8.0 arc min and between 4.0 and 12.0 arc min ($p < 0.05$) but not between 8.0 and 12.0 arc min.

Figure 3 also shows that the sign of the mean bias of the PSE was positive for every disparity and every number of elements in both fixation conditions, indicating that the total-element overestimation phenomenon was experienced by the observers in this experiment. As in Experiment 1, this finding is consistent with the results of the three-way mixed-design ANOVA, which showed that the intercept was significantly different from zero, $F(1, 14) = 62.84$, $p < 0.001$. In other words, this statistical significance and the positive sign of the smallest value of the mean bias indicate that the stereoscopic 3-D stimulus with two overlaid surfaces was perceived to contain more elements than the stereoscopic 2-D stimulus.

Discussion

The results showed that binocular disparity had an effect on the total-element overestimation phenomenon

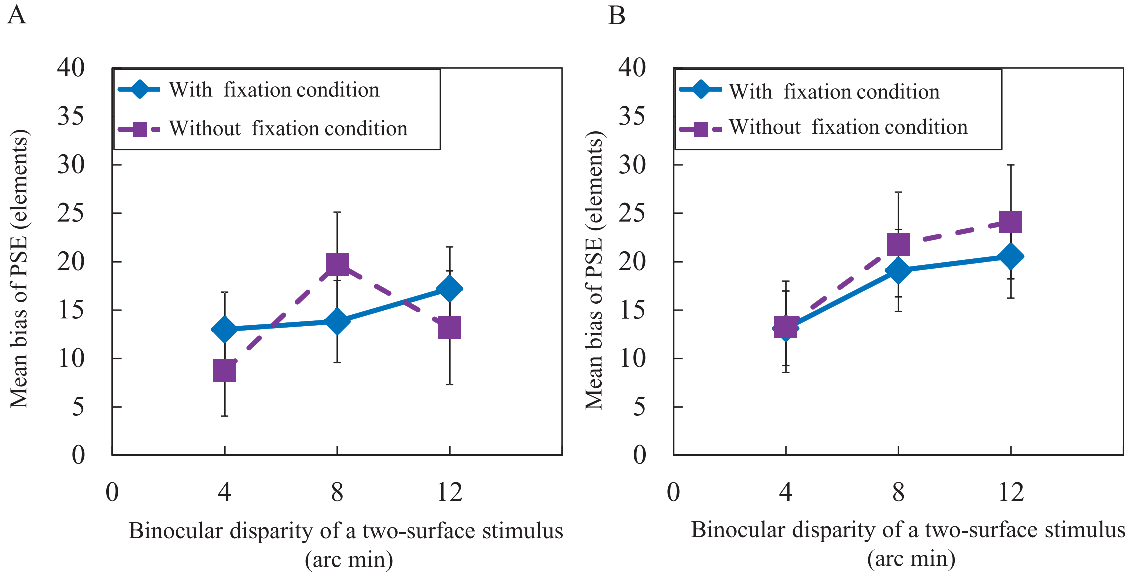

Figure 3. Results from Experiment 2. Mean biases of the PSE as a function of binocular disparity of a two-surface stimulus with (A) 150 elements and (B) 300 elements. The diamond symbols indicate the extent of the bias for the with-fixation condition, and the square symbols indicate the extent of the bias for the without-fixation condition. Error bars are ±1 *SE*.

and that manipulation of fixation did not. The first result showing the effect of binocular disparity apparently contradicts with Schütz's (2012) Experiment 3, where the number of elements in the back surface was overestimated compared to that of the front surface and the degree of overestimation was constant irrespective of the binocular disparity between the two surfaces. This difference in findings can be due to the difference between the tasks used in this experiment and in Schütz's. In the present experiment, observers were asked to judge the numerosity of elements that were in a 2-D or a 3-D stimulus as a whole, while in Schütz's, observers were asked to compare the number of elements in the front surface against that of a back surface of a 3-D stimulus with two stereo motion-transparent surfaces. Furthermore, the range of disparities that were used varied between the two studies. Specifically, the disparities between the two stereo motion-transparent surfaces (3-D stimulus with two-surfaces) were 3.4, 11.24, and 25.7 arc min in Schütz's study, and the disparities used in the present study were 4.0, 8.0, and 12.0 arc min. In short, our findings do not necessarily contradict with those reported by Schütz.

Based on the findings in the present experiment, we can deduce that the total-element overestimation phenomenon occurs at the level of visual processing at which representation of surfaces in depth is achieved. Note that neither the presence nor absence of a fixation stimulus had an effect on the total-element overestimation phenomenon, which suggests that vergence eye movement is unlikely to have played a role in the phenomenon; with controlled fixation, double vision that could have arisen because of extreme vergence eye positions would not have occurred. Consequently, we can safely argue that, because both a 2-D stimulus and a 3-D stimulus project the same number of elements in each eye, the difference in numerosity judgments between the two stimuli may relate to the fact that the 3-D stimulus produces relative depth perception but the 2-D stimulus does not.

As discussed earlier, Schütz (2012) reported a phenomenon that when numerosity judgments of a 3-D stimulus were made separately for each surface, the number of elements in the back surface is overestimated and that of the front surface is slightly underestimated. This reported asymmetry in the numerosity judgment of elements between front and back surfaces suggests that the total-element overestimation phenomenon reported in the present study may be attributed to an overestimation of the number elements in the back surface. We examined this suggestion in Experiment 3.

Experiment 3

In Experiment 3, we examined whether or not the total-element overestimation phenomenon for a 3-D stimulus with two-surfaces reported in Experiments 1 and 2 is due to an increase in the perceived number of elements in its back surface (i.e., that it can be attributed to the back-surface-element overestimation phenomenon). If there is an overestimation of the number of elements for the back surface and if the perceived number of elements in each surface of a two-surface stimulus is being summed up in judging the

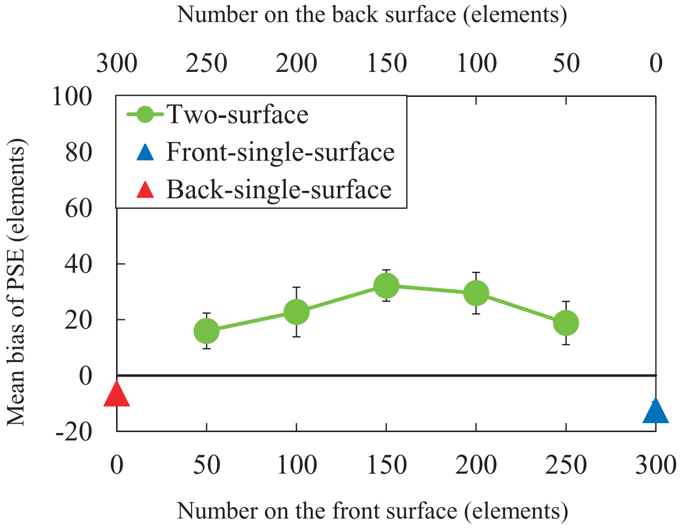

Figure 4. Results from Experiment 3. Mean biases of the PSE as a function of the number of elements in the front surface of a two-surface stimulus, with the total number of elements in the whole stimulus kept constant at 300 elements. Green circles indicate the extent of the bias for the two-surface condition. The blue triangle symbol indicates the extent of the bias for the front-single-surface with disparity condition, and the red triangle symbol indicates the extent of the bias for the back-single-surface with disparity condition. Error bars are ±1 *SE*.

total number of elements in the stimulus, then the back-surface-element overestimation phenomenon can explain the total-element overestimation phenomenon. We examined this idea by using the prediction derived from a set of equations that was proposed by Schütz (2012) to describe the back-surface-element overestimation phenomenon (see Appendix). According to the equations, when the physical total number of elements is kept constant, it is predicted that the perceived total number of elements will increase when either the physical number of elements in the front is increased or the back is decreased (see Equation A3 in Appendix).

Method

Stimuli

The standard was either a two-surface stimulus with the surfaces located in depth and straddling the monitor plane, or a single-surface stimulus that was either located in the front (crossed disparity) or in the back (uncrossed disparity) behind the monitor plane (see Figure 1B). The total number of elements for the standard, either single-surface or two-surface, was kept constant at 300 elements. For the two-surface standard, the ratio (x:y) of the number of elements in its front surface (x) and those in its back surface (y) were manipulated from 50:250 to 250:50, with increments and decrements of a fixed step size of 50 elements. The total disparity (interplane disparities summed up) of the two-surface standards was 12.0 arc min, and the stereoscopic element pairs used to generate the single front and back surfaces had 6.0 arc min crossed and uncrossed disparities, respectively, with respect to the monitor plane. The comparison was a middle-single-surface stimulus with zero disparity (see Figure 1B). The number of elements on the comparison was varied from 220 to 380 elements using an incremental step size of 40 elements for each of the 5 two-surface standards and each of the front-single-surface and back-single-surface standards.

Procedure

The experiment was conducted in one session with seven blocks of trials. Each block was for each of the seven standards: the two-surface standard with five different ratios of elements in its front surface to that of its back surface, the front-single-surface standard and the back-single-surface standard. The presentation order of the standards was randomized for each of the eight observers. In each block, the comparison was presented five times at each of two different presentation locations (right or left). Thus, in total there were 350 trials (7 standards × 5 numbers of elements on the comparison × 2 locations × 5 repetitions) for each observer.

Results

We analyzed separately the PSE for the two-surface standards and the single-surface standards. First, we conducted a one-way repeated measures ANOVA (5 two-surface standards) on the bias of the PSE of the two-surface standards. The analyses showed that the main effect of the two-surface standards was not statistically significant. This result is reflected in Figure 4, which shows the mean bias of the PSE as a function of the number of elements in the front surface of the standard. As can be seen in Figure 4, the mean biases of the 5 two-surface standards are fairly constant.

Figure 4 also shows that the sign of the mean bias of the PSE is positive for each of the five two-surface standards, indicating that the total-element overestimation phenomenon was observed for the two-surface stimuli whose ratio of the elements in the front and back surfaces had elements ranging from 50:250 to 250:50. As in Experiments 1 and 2, this observation is consistent with the result of an intercept test in the one-way ANOVA, described above, which indicated that the intercept was significantly different from zero, $F(1, 7) = 17.95$, $p < 0.001$. This significance and the positive sign of the smallest value of the mean bias indicate that the stereoscopic 3-D stimulus was perceived to contain more elements than the stereoscopic 2-D stimulus within the range of ratios (of the number of elements in the front surface to that in the back surface) that were used in the present experiment.

Second, we compared the mean bias (−12.3) of the PSE for the front-single-surface standard with that (−6.4) for the back-single-surface standard. A t test showed that there was no statistical difference in the bias between the two conditions. Note the triangular symbols indicated on the abscissa in Figure 4; at 0 elements, observers viewed the back-single-surface standard only, and at 300 elements, observers viewed the front-single-surface standard only. Importantly, the mean biases of the front-single-surface and back-single-surface standards are not that different from each other. Furthermore, a negative value of the mean bias in either standard suggests that the overestimation phenomenon was not observed for the 2-D single-surface stimuli that were located at different front-parallel planes. A t test showed that the bias of the PSE was significantly different from zero for the front-single-surface stimulus, $t(7) = 4.27$, $p < 0.01$, while not for the back-single-surface stimulus, suggesting that the perceived number of elements in the front-single surface was underestimated, compared to the physical number of elements in the middle-single surface.

Discussion

The result that there was no statistically significant difference in the bias of the PSE between any pair of the five sets of elements on the two-surface standard is inconsistent with the prediction that when the total physical number of elements on the two-surface standards is kept constant, increasing the physical number of elements on the front surface would result in an increase in the total perceived number of elements on the two-surface stimulus (see Appendix). As can be seen in Figure 4, the total perceived number of elements remained constant as a function of the physical number of elements on the front surface.

There are two possible explanations for this apparent difference between the present results and the prediction. One is that the overestimation phenomenon reported in this study may not be explained by an overestimation of perceived elements in the back surface. The other is that the overestimation of the back surface might be observed only for moving elements, as was the case with Schütz (2012), and not for static elements as was the case in this experiment. (Note that the prediction was based upon the result of Schütz's experiment 7.) We investigated which explanation is more plausible in Experiment 4.

As a final point, the result that there was no statistically significant difference in the bias of the PSE between the front-single surface and back-single-surface stimuli supports our discussion in Experiment 1 that stereoscopic size constancy may not have played a role in the total-element overestimation phenomenon reported in this study. As discussed, according to the principle of stereoscopic size constancy, the perceived size of the elements in the back-single-surface may be perceived to be larger than that in the front-single-surface. The fact that the perceived number of elements was the same between the front and back surfaces in this experiment suggests that the stereoscopic size constancy is not an important factor in the numerosity judgment for the stimuli we used. Accordingly, it is difficult to explain the total-element overestimation phenomenon in terms of stereoscopic size constancy. (Note that the perceived number [288] in the front-single-surface was significantly smaller than the physical number [300] in the middle-single-surface and thus, this result is not inconsistent with stereoscopic size constancy. However, the result showed that the underestimation of the perceived number of elements is inconsistent with the overestimation phenomenon reported in this study.)

Experiment 4

In Experiment 4, we examined whether the difference between the results of Experiment 3 and the prediction based on the results of Schütz (2012) is explainable by the difference between the tasks used or the types of stimulus used in this study and that of Schütz. If an overestimation of the number of elements in the back surface and an underestimation of the elements in the front surface are observed for a static 3-D stimulus (as well as a moving stimulus, such as in Schütz's study), then the difference between the results and the prediction may be based on the difference in the task. If this does not occur, then the difference may be based on the type of stimulus used. To examine which explanation is more plausible, the perceived numbers of the dot elements in the front back surfaces were measured separately.

Method

Stimuli

The standard stimuli used in this experiment were the same as those used in Experiment 3: five different two-surface stimuli, one front-surface stimulus, and one back-surface stimulus (see Figure 1B). When the perceived number of elements in the front surface was judged, the standard was chosen from the five different

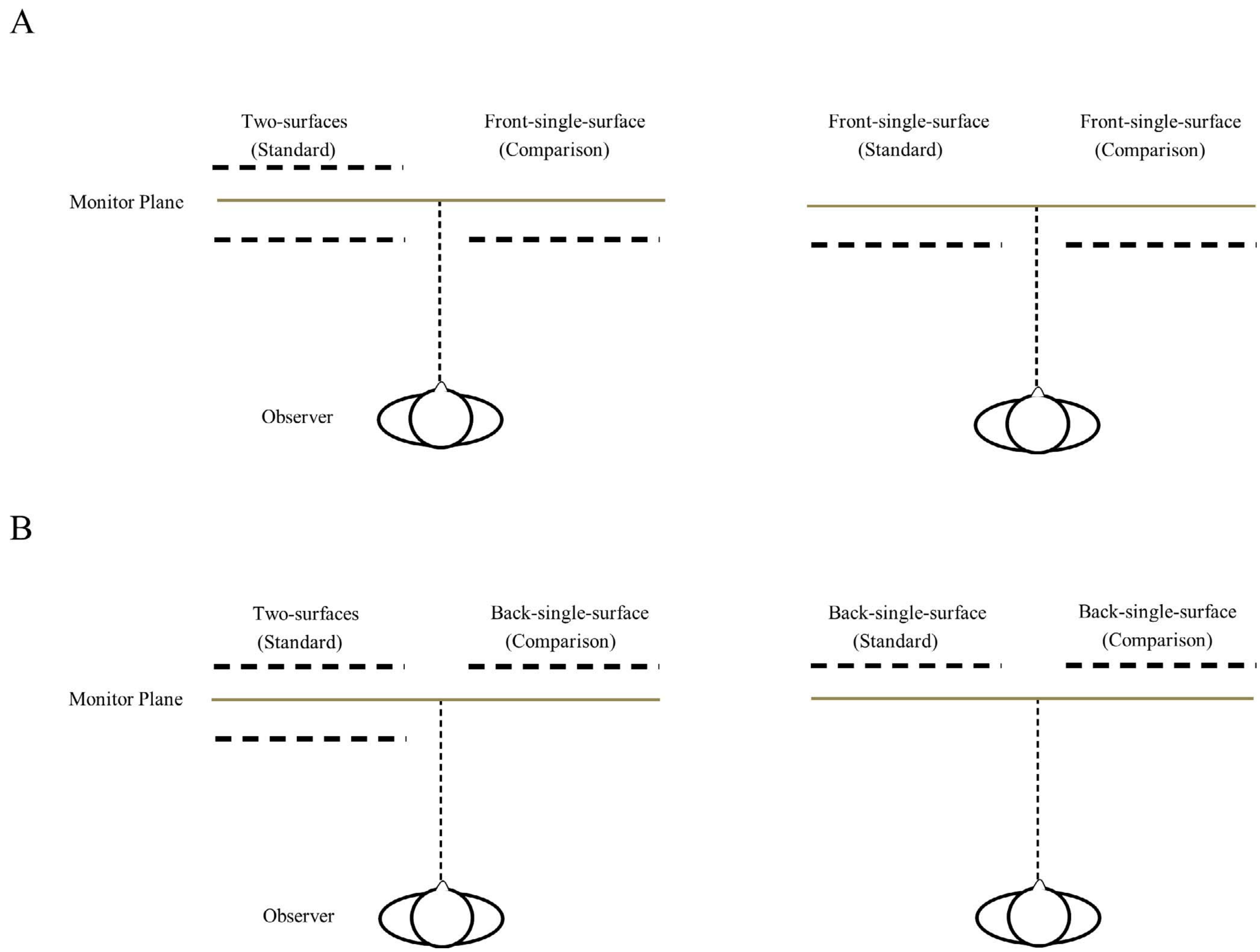

Figure 5. Schematic top view illustration of observers' perception of the stimuli for (A) the front judged surface, and (B) the back judged surface conditions. In both A and B, the top views on the left illustrate the two-surface standard and the single-surface comparison that were presented side by side; the top views on the right illustrate the single-surface standard and the single-surface comparison.

two-surface stimuli and the front-single-surface stimulus (Figure 5A). When the perceived number of elements in the back surface was judged, the standard was chosen from the same five different two-surface stimuli and the back-single-surface stimulus (Figure 5B). Thus, six different standards were used for each of the two judged-surface (front judged-surface and back judged-surface) conditions. The front and back surfaces of the two-surface standard, with a total of 300 elements, had a variable number of elements ranging from a front-to-back ratio of 50:250 to a ratio of 250:50, with an incremental step size of 50 elements. The front-single-surface standard had 300 elements, and the back-single-surface standard also had 300 elements.

The comparison stimulus used in this stimulus was the front-single-surface or the back-single-surface stimulus (see Figure 1B). For the front-judged-surface (or the back-judged-surface) condition of the 5 two-surface standards that had 50, 100, 150, 200, and 250 elements, the number of elements on the comparison was varied at five levels with the middle level set to the same number of elements as the standard that was being tested. For different observers we used different initial values and ranges of the step sizes for the five levels used for the comparison, because the ranges were adjusted to match the responses necessary to fit the psychometric function for each observer. The initial values ranged from 10 to 240, and the step sizes ranged from 10 to 40. When the observers were asked to judge the number of elements in the front- and back-single-surface standards, the comparison was also at the same frontal-parallel plane as the front-single-surface and back-single-surface standards, respectively (see Figure 5). In this condition, the number of elements on the comparison was varied from 240 to 360 elements in increments of a step size of 30 elements for each of the front-single-surface and back-single-surface stimuli.

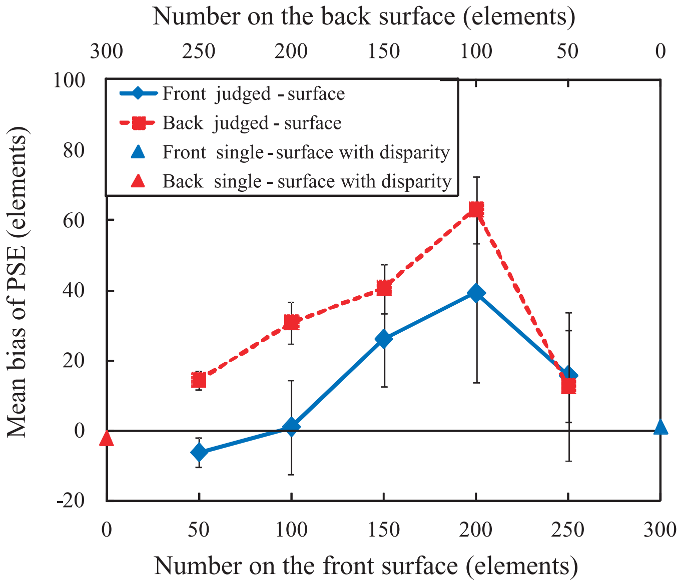

Figure 6. Results from Experiment 4. Mean bias of the PSE as a function of the number of elements in the front surface of a two-surface stimulus, with the total number of elements in the whole stimulus kept constant at 300 elements. The blue diamond and red square symbols indicate the extent of the bias for estimates of the number of elements in the front surface and in the back surface, respectively, for the two-surface condition standard. The blue triangle symbol indicates the extent of the bias for the front-single-surface with disparity condition, and the red triangle symbol indicates the extent of the bias for the back-single-surface with disparity condition. Error bars are ±1 *SE*.

Procedure

There were two experimental sessions: one in which numerical judgments of the front surface were required and the other in which numerical judgments for the back surface were required. The order of the sessions (front-judged-surface and back-judged-surface sessions) was randomized for each observer. Each judged-surface session consisted of six blocks of trials, and each block was for each of the six different standards. The presentation order for the six standards was randomized for each observer. In each block, the comparison was presented five times either on the right or the left of the standard. Thus, in total there were 600 trials (2 judged-surfaces × 6 standards × 2 locations × 5 repetitions × 5 number of elements on the comparison) for each observer.

Results

We analyzed the bias of the PSE for the 5 two-surface standards and two single-surface standards, separately, as in Experiment 3. First, we conducted a two-way repeated measures ANOVA (2 judged-surfaces × 5 two-surface standards) on the bias of the PSE for the two-surface standard. The analyses showed that the main effects of judged surfaces, $F(1, 6) = 7.86$, $p < 0.05$, and the number of elements, $F(4, 24) = 5.83$, $p < 0.01$, were statistically significant, but their interaction was not. The significant main effects of the judged surfaces and the number of elements can be seen in Figure 6. It shows the mean bias of the PSE as a function of the number of elements in the front surface of the two-surface standard either for the front or back judged surface. Post hoc analyses (Tukey tests) showed that the differences of the mean biases between 50 and 200 elements, 100 and 200 elements, and 200 and 250 elements were statistically significant ($p < 0.05$), but not for other comparisons. This result is consistent with Figure 6, which shows that the mean bias increases with the number of elements in the front surface except for the 250-element condition.

Figure 6 also shows that the sign of the mean bias of the PSE for the two-surface standards is different among the five standards in the front judged surface and the same (positive) among them in the back judged surface. We performed the intercept test in a one-way repeated measures ANOVA, separately for the front and back judged surfaces to examine whether or not the smallest value of the mean bias was different from zero, as in Experiments 1, 2, and 3. The ANOVA for the back judged surface showed that the intercept was significantly different from zero, $F(1, 6) = 12.50$, $p = 0.012$, suggesting that an overestimation of elements in the back surface was observed in all the 5 two-surface standards. On the other hand, the ANOVA for the front judged surface showed no significant difference, suggesting that the smallest mean bias of the PSE was neither overestimated nor underestimated. Accordingly, the result of post hoc analyses on the mean biases among the 5 two-standards described in the previous paragraph suggests that the perceived number of elements at the 200 elements in the front judged surface is larger than zero or overestimated.

Finally, we compared the mean bias (1.10) of the PSE for the front-single-surface standard with that (−2.17) for the back-single-surface standard. The two biases of PSE are plotted in Figure 6 (blue triangle and red triangle symbols, respectively). This result indicates that when two stimuli (comparison and standard) presented side by side were seen at the same depth plane, the perceived number of elements for each stimuli appeared to be the same.

Discussion

Comparing the results of this experiment to those of experiment 7 of Schütz (2012) indicates that they are consistent to a certain extent. The overestimation of the elements in the back surface was observed in both studies. On the other hand, a relatively small amount of underestimation was observed for the front surface in Schütz's study but not in this current experiment. That is, the overestimation of elements in a back surface (i.e., the back-surface-element overestimation phenomenon) can be seen in both types of 3-D stimulus: one consisting of dots moving in two different directions and the other consisting of static random elements. With respect to the motivation for conducting this particular experiment, this outcome suggests that the difference between the results of Experiment 3 in this study and in Schütz's study may have arisen from a difference between the tasks rather than from a difference in the stimulus types used.

Note, however, the results of Experiment 3 are difficult to explain in terms of the overestimation of the back surface elements only, although the back-surface-overestimation phenomenon was observed in Experiment 4. If the visual system sums the perceived number of elements for each surface of the two-surface stimulus in judging the total number of elements, the perceived total number of elements would be similar to the summation between the perceived number of elements for the front surface and that for the back surface. However, the summation of the PSE bias of the front surface and that of the back surface obtained in Experiment 4 is more than twice as large as that obtained in Experiment 3, in particular when the front surface contained 100, 150, and 200 elements. To explain the results of Experiment 3 in terms of the results of Experiment 4, we need to assume that the visual system sums the perceived numbers for a front surface and back surface nonlinearly so that the summed numerosity is constant as in Experiment 3.

As the final point, we discuss the sudden drop in the perceived numerosity of the back judged surface at the 250 elements condition in the front surface (see Figure 6). In this condition, the number (or density) of the front surface is 5 times larger than that of the back surface. This may suggest that when the density of the back surface is much smaller than that of the front surface, the back-surface-overestimation phenomenon is weakened. This interpretation of the result is consistent with the back-surface bias hypothesis (Tsirlin et al., 2012), which assumes that a process assigns the dots in the back surface to their surrounding blank areas. If the process is less likely to operate when the element density is low, a sudden drop can be explained.

General discussion

In this study we found that a stereoscopic 3-D stimulus, with elements that are distributed over one or more of its surfaces, is perceived to contain more elements than a stereoscopic 2-D stimulus, even when the two stimuli contain the same number of elements and produce the same number of elements on the retinae. Experiment 1 showed that a 3-D stimulus with either two or three surfaces is perceived to contain more elements than a 2-D stimulus with a single-surface, for a relatively large range of numbers of elements (72–600 elements). Experiment 1 also showed that the 3-D stimulus with three surfaces is perceived to contain more elements than that with two stereo-surfaces. Experiment 2 showed that the overestimation of the total number of elements for a 3-D stimulus with two surfaces depends on its binocular disparity, particularly when it is less than 12.0 min arc. The experiment also showed that the overestimation does not depend on observers' eye positions. Experiment 3 showed that the distribution of elements between the front and back surfaces of a 3-D stimulus with 300 elements has no effect on the overestimation of the total number of element of the stimulus. Experiment 4, however, showed that the distribution has an effect on the overestimation of the number of elements on the back surface and not on the estimation of the number of elements on the front surface. These results indicate that the numerosity judgment for a 3-D stimulus is overestimated compared to that in a 2-D stimulus and suggest that the overestimation of the total number of elements for a 3-D stimulus with two surfaces is not referable to an overestimation of elements in its back surface.

Although as discussed in Experiment 4, the total-element overestimation phenomenon cannot be simply due to the overestimation of the elements in the back surface, the back-surface bias hypothesis—when two surfaces overlap in depth, the back surface affects numerosity judgment (Schütz, 2012)—is still attractive. As discussed in the Introduction, the hypothesis assumes that the dots in the back surface are assigned to their surrounding blank areas to form an opaque background surface through a higher order process, making its dot density appear denser than that of the front surface (Tsirlin et al., 2012). If a surface with a denser number of dots can lead to the perception of a greater number of dot elements, the hypothesis can explain why the overestimation phenomenon was observed in this study and in Schütz (2012) but not in Bell et al. (2015). As discussed in Experiment 1, the difference between the former and the latter can be due to the difference in the stimuli used between them; the stimuli Schütz used contained the back surface but not Bell et al.'s. If the back surface plays an important role in both the total-element overestimation and the back-surface-element overestimation phenomena, the fact that the overestimation was not observed in Bell et al. is explainable.

The total-element overestimation phenomenon can also be accounted for by a hypothesis that assumes the visual system takes into account elements occluded by ones located in the front surface of a 3-D stimulus in estimating the number of elements contained as a whole. This hypothesis can be deduced from the fact that a stereo-transparent surface often simulates a situation where nontransparent objects on a front surface of a 3-D stimulus occlude or hide objects that are behind (Aida et al., 2015; Tsirlin et al., 2008, 2012), and thus, there can be "unseen" elements occluded completely by ones that are on the front surface. If the visual system takes into account this possibility, the system may add an additional number of unseen elements. If this were the case, the total number of perceived elements in a 3-D stimulus would appear more numerous than that of a 2-D stimulus.[4] This occlusion hypothesis is consistent with the result obtained in Experiment 1 in that the perceived elements for a three-surface stimulus was larger than those for a two-surface stimulus when the two stimuli had the same physical number of elements. If the surface number increases, the number of potentially occluded elements will also increase. However, this hypothesis is inconsistent with the result of Experiment 3. Manipulation of the number of elements in the front surface of a 3-D stimulus did not affect the numerosity overestimation when the total number of elements in a 3-D stimulus with two surfaces was kept constant, whereas it did have an effect when observers were asked to estimate and report the number of elements in each surface in Experiment 4. Thus, the occlusion hypothesis cannot explain the complete results of this study.

Finally, we reported a phenomenon that when observers are asked to compare the number of elements in a 3-D (two-surface) stimulus and that of a 2-D (single-surface) stimulus, both of which contain the same number of elements, the observers would judge the number of elements of the former stimulus to be greater than that of the latter. We discussed two hypotheses that could explain the overestimation. One hypothesis assumes that the back surface of the two-surface stimulus affects the numerosity or density judgment through a higher order process that makes the back surface appear more dense (Tsirlin et al., 2012). The other assumes that the visual system takes into account elements occluded by those in the front surface of a two-surface object. Both ideas, however, have difficulty explaining the whole set of results obtained in this study and in the literature, suggesting that numerosity overestimation of a 3-D stimulus may be the result of several processes.

Keywords: numerosity, binocular stereopsis, stereo-transparency, binocular disparity

Acknowledgments

This work was partly supported by Grants-in-Aid for Scientific Research (25-10777, H03463, and 23330215) provided by the Japanese Ministry of Education, Science, and Culture. Part of this study was reported in Vision Sciences Society (2013) held at Naples. The third author, Koichi Shimono, worked on this study while he was at York University, Toronto, Canada, as a Japan Society for the Promotion of Science Research Exchange Fellow from Tokyo University of Marine Science & Technology. The authors would like to thank Linda Lillakas and Dr. Hiroshi Ono for their helpful comments on the manuscript. The authors also thank Satoshi Shimono for programming the experiments.

Commercial relationships: none.
Corresponding author: Saori Aida.
Email: saori.aida.t@gmail.com.
Address: Graduate School of Marine Science and Technology, Tokyo University of Marine Science and Technology, Koto-ku, Tokyo, Japan.

Footnotes

[1] Note that although in this study we use the term *numerosity judgment*, the term does not necessarily imply that the judgment is mediated through a mechanism for numerical sense. Based on studies of numerosity judgments for a 2-D stimulus (Burr & Ross, 2008; Dakin et al., 2011; Durgin, 2008; Ross & Burr, 2010), it is not known yet whether the numerosity judgment performed by observers in this study is mediated through a mechanism for numerical sense or one for density sense, or whether the judgment is mediated through a mechanism for both senses. Because the size of the 2-D and 3-D stimuli used in each experiment were identical, the terms *numerosity* and *density* are interchangeable. However, we use the numerosity judgment term because we asked observers to make numerosity judgments.

[2] In a number of studies (e.g., Krueger, 1972; Indow & Ida, 1977), dot elements were distributed in different depths on a slanted surface. For example, in Krueger (1972), dot elements were depicted on a piece of white paper that was placed on a table and viewed at an angle from above by observers who were standing by the table. However, we do not regard the stimulus used by Krueger (1972) as a 3-D stimulus, because dot elements were placed on a single surface. In addition, note that the results of Krueger (1972) are consistent with those obtained using a frontal flat surface (see Krueger, 1984).

[3] One might think that the phenomenon is due to cross-talk between the right-eye and left-eye views, which can arise because of the anaglyph method that was used. When the image intended for one eye leaks to the other eye because of cross-talk, observers will see more elements than what each eye would see. If this happened, then an overestimation of the number of elements can be attributed to cross-talk. However, this explanation is not likely because in Experiment 3 we found that the single surface containing 300 dot elements with zero disparity, which surely produced 300 elements on the retina of each eye, was judged to contain the same or more elements than the single surface with cross or uncrossed disparity (see Figure 4). This result indicates that cross-talk, if any, did not increase the number of the perceived elements for each surface, suggesting that cross-talk most likely did not cause the overestimation results found in Experiments 1 and 2. Furthermore, in a supplementary experiment not reported here, we found the overestimation phenomenon using a mirror stereoscope, which presented a separate image to each eye, suggesting that the overestimation phenomenon can be observed using either type of stereoscope and is not specific to the anaglyph method.

[4] Recently, Zeiner, Spitschan, and Harris (2014) found that dot elements presented on a stereoscopic surface behind a surrounding reference were perceived to be more numerous when they were occluded by horizontally oriented bars, which were perceived at the same depth plane as the reference (see their figure 5).

References

Aida, S., Kusano, T., & Shimono, K. (2013). Overestimation of the numbers of elements in a three-dimensional stimulus compared with a two-dimensional stimulus. *Journal of Vision*, *13*(9):1229, doi:10.1167/13.9.1229. [Abstract]

Aida, S., Shimono, K., & Tam, W. J. (2015). Magnitude of perceived depth of multi stereo-transparent-surfaces. *Attention, Perception, & Psychophysics*, *77*(1), 190–206, http://www.ncbi.nlm.nih.gov/pubmed/25120178, doi:10.3758/s13414-014-0746-8.

Akerstrom, R. A., & Todd, J. T. (1988). The perception of stereoscopic transparency. *Perception & Psychophysics*, *44*, 421–432, http://www.ncbi.nlm.nih.gov/pubmed/3226891, doi:10.3758/BF03210426.

Allik, J., Tuulmets, T., & Vos, P. G. (1991). Size invariance in visual number discrimination. *Psychological Research*, *53*, 290–295, http://www.ncbi.nlm.nih.gov/pubmed/1792300, doi:10.1007/BF00920482.

Bell, J., Manson, A., Edwards, M., & Meso, A. I. (2015). Numerosity and density judgments: Biases for area but not for volume. *Journal of Vision*, *15*(2):18, 1–14, doi:10.1167/15.2.18. [PubMed] [Article]

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology*, *18*(6), 425–428, http://www.ncbi.nlm.nih.gov/pubmed/18342507, doi:10.1016/j.cub.2008.02.052.

Burt, P., & Julesz, B. (1980). Modifications of the classical notion of Panum's fusional area. *Perception*, *9*, 671–682, http://www.ncbi.nlm.nih.gov/pubmed/7220240, 10.1068/p090671.

Dakin, S., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences, USA*, *108*(49), 19552–19557, http://www.ncbi.nlm.nih.gov/pubmed/22106276, doi:10.1073/pnas.1113195108.

Durgin, F. (2008). Texture density adaptation and visual number revisited. *Current Biology*, *18*, R855–R856, http://www.ncbi.nlm.nih.gov/pubmed/18812077, doi:10.1016/j.cub.2008.07.053.

Ginsburg, N., & Nicholls, A. (1988). Perceived numerosity as a function of item size. *Perceptual & Motor Skills*, *67*(2), 656–658, http://www.ncbi.nlm.nih.gov/pubmed/3217220, doi:10.2466/pms.1988.67.2.656.

Howard, I. P., & Rogers, B. (2012). *Perceiving in depth, Volume 2: Stereoscopic vision*. Oxford: Oxford University Press.

Indow, T., & Ida, M. (1977). Scaling of dot numerosity. *Perception & Psychophysics*, *22*(3), 265–276, doi:10.3758/BF03199689.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, IL: University of Chicago Press.

Krueger, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, *11*, 5–9, doi:10.3758/BF03212674.

Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & Psychophysics*, *35*, 536–542, doi:10.3758/BF03205949.

Oyama, T. (1974). Perceived size and perceived distance in stereoscopic vision and an analysis of their causal relations. *Perception & Psychophysics*, *16*(1), 175–181, doi:10.3758/BF03203271.

Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision*, *10*(2):10, 1–8, doi: 10.1167/10.2.10. [PubMed] [Article]

Schütz, A. C. (2012). There's more behind it: Perceived depth order biases perceived numerosity/density. *Journal of Vision*, *12*(12):9, 1–16, doi:10.1167/12.12.9. [PubMed] [Article]

Tibber, M. S., Greenwood, J. A., & Dakin, S. C. (2012). Number and density discrimination rely on a common metric: Similar psychophysical effects of size, contrast, and divided attention. *Journal of Vision*, *12*(6):8, 1–19, doi:10.1167/12.6.8. [PubMed] [Article]

Tokita, M., & Ishiguchi, A. (2010). How might the discrepancy in the effects of perceptual variables on numerosity judgment be reconciled? *Attention, Perception, & Psychophysics*, *72*(7), 1839–1853, http://www.ncbi.nlm.nih.gov/pubmed/20952782, doi:10.3758/APP.72.7.1839.

Tsirlin, I., Allison, R. S., & Wilcox, L. M. (2008). Stereoscopic transparency: Constraints on the perception of multiple surfaces. *Journal of Vision*, *8*(5):5, 1–10, doi:10.1167/8.5.5. [PubMed] [Article]

Tsirlin, I., Allison, R. S., & Wilcox, L. M. (2012). Perceptual asymmetry reveals neural substrates underlying stereoscopic transparency. *Vision Research*, *54*, 1–11, http://www.ncbi.nlm.nih.gov/pubmed/22186226, doi:10.1016/j.visres.2011.11.013.

Valsecchi, M., Toscani, M., & Gegenfurtner, K. R. (2013). Perceived numerosity is reduced in peripheral vision. *Journal of Vision*, *13*(13):7, 1–16, doi:10.1167/13.13.7. [PubMed] [Article]

Yeh, Y. Y., & Silverstein, L. D. (1990). Limits of fusion and depth judgment in stereoscopic color displays. *Human Factors*, *32*(1), 45–60, http://www.ncbi.nlm.nih.gov/pubmed/2376407, doi:10.1177/001872089003200104.

Zeiner, K. M., Spitschan, M., & Harris, J. M. (2014). Perceptual integration across natural monocular regions. *Journal of Vision*, *14*(3):5, 1–14, doi:10.1167/14.3.5. [PubMed] [Article]

Appendix

Details of prediction in Experiment 3

Based upon the results of his experiment 7, Schütz (2012) proposed that the perceived number on elements in each surface of a 3-D stimulus with two surfaces can be described as a linear function as follows,

$$N_{f'} = N_f - 0.1N_b, \tag{A1}$$

and

$$N_{b'} = N_b + 0.2N_f, \tag{A2}$$

where $N_{f'}$ and $N_{b'}$ are the perceived number of elements for the front and the back surface, respectively, N_f and N_b are the physical number of elements for the front and the back surface, respectively. From the two equations, we obtain,

$$N_{f'} + N_{b'} = N_f + N_b + 0.1(2N_f - N_b), \tag{A3}$$

where $N_{f'} + N_{b'}$ and $N_f + N_b$ are the total perceived number of elements and the total physical number of elements, respectively. In Experiment 3, we increased the physical number of elements on the front surface and decreased that on the back surface while keeping the whole number of elements on the two surfaces constant. Thus, Equation A3 predicts that the perceived total number of elements will increase as the physical elements on the front surface increases.