How might the discrepancy in the effects of perceptual variables on numerosity judgment be reconciled?

MIDORI TOKITA AND AKIRA ISHIGUCHI
Ochanomizu University, Tokyo, Japan

It has been claimed that a genuinely abstract numerical representation would be capable of representing the numerosity of any set of discrete elements independently of the attributes of the individual elements. In practice, however, perceptual variables, such as element size, density, and others, affect numerosity judgment. In this study, we examined how perceptual variables affect the performance of numerosity judgments under the control of factors such as amount of practice and set size. In Experiment 1, we introduced the numerosity comparison task to less experienced observers in order to examine the effect of element size and array area. In Experiment 2, we examined whether and how practice would influence precision and accuracy in numerosity comparison and demonstrated that the effects of perceptual variables mostly disappeared. In Experiment 3, we examined whether the effect of practice could transfer to the performance in different stimulus conditions. Taking the results together, we demonstrate that differences in practice might be the source of inconsistent results for perceptual variables.

The ability to judge the relative numerosity of items in the visual field is quite adaptive. For various species, many situations require them to judge which items are greater in number without counting serially. For instance, when the decision is made to fight or run, it is necessary to judge which side, friends or foes, is more numerous. Alternatively, when collecting food, it is crucial to judge which location has more pieces of fruit or a greater abundance of prey.

Substantial empirical evidence has supported the idea that humans possess innate neural mechanisms that generate approximate, not precise, numerical representations. Results from studies of numerical competence in infants, young children, and nonhuman animals have shown that the approximate numerical representation system is evolutionally old and is equipped early in human development (e.g., Brannon, 2006; Cantlon, Platt, & Brannon, 2009; Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004; Whalen, Gallistel, & Gelman, 1999). Futhermore, converging empirical findings from several areas of cognitive neurosience have argued for biologically determined mechanisms for approximate numerical representation (e.g., Cantlon, Brannon, Carter, & Pelphrey, 2006; Cappelletti, Barth, Fregni, Spelke, & Pascual-Leone, 2007; Dehaene, Dehaene-Lambertz, & Cohen, 1998; Eger, Sterzer, Russ, Giraud, & Kleinschmidt, 2003; Nieder & Dehaene, 2009; Nieder, Freedman, & Miller, 2002; Nieder & Miller, 2003; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). For example, neurophysiological research has provided evidence for the existence in monkeys' pre-frontal cortex of neurons that are tuned for small numbers (Nieder & Miller, 2003) and large numbers such as 30 (Nieder & Merten, 2007). The characteristics of the neurons suggested that the performance of discrimination obeys the Weber law, which implies that numerical tuning is approximate and is broader for larger numerosities. Many studies of numerosity discrimination have shown that discriminability depends on the ratio of the numerosities to be compared (e.g., Allik, Tuulmets, & Vos, 1991; Brannon, 2006; Krueger, 1972, 1984; Nieder & Dehaene, 2009). Piazza et al. found that the intraparietal site coding for number in humans is compatible with that observed in macaque monkeys. It should be noted that in this study, the approximate numerical representation referred mainly to the representation of nonverbal or nonsymbolic numerical value.

In line with these arguments, it has been claimed that a genuinely abstract numerical representation would be capable of representing the numerosity of any set of discrete elements, whether they were events or items, presented sequentially or simultaneously, independently both of the spatial arrangement of the stimuli and of the attributes of the individual elements such as size, color, and shape. In practice, however, perceptual variables, such as element size, density, and others, affect numerosity judgments, and this issue is still a matter of debate. Several studies have asserted the influence of perceptual variables in numerical estimation and comparison; the dissociation between physical numerosity and perceived numerosity has been reported in extensive investigations in adults (e.g., Allik

M. Tokita, tokita.midori@ocha.ac.jp

& Tuulmets, 1991; Ginsburg, 1991; Ginsburg & Nicholls, 1988; Krueger, 1972; Shuman & Spelke, 2006; Sophian & Chu, 2008), as well as in young children (Clearfield & Mix, 1999; Cordes & Brannon, 2008, 2009; Mix, Huttenlocher, & Levine, 2002; Rousselle, Palmers, & Noel, 2004).

Among several perceptual variables, studies of the effects of element size and display size of the array area have not provided conclusive evidence. Different approaches have yielded contradictory results. For example, Krueger (1972) demonstrated that arrays containing 25–200 dots were estimated to be less numerous when they were more densely spaced than when they were spread over a larger area. Similarly, Hollingsworth, Simmons, Coates, and Cross (1991) demonstrated that adults' tendency to underestimate larger numerosities was lessened when the elements were more spread out, so that denser arrays elicited lower estimates of numerosity in their research. On the other hand, in a comparative judgment study, Allik et al. (1991) demonstrated that perceived numerosity was not affected by the size of the array within which elements were distributed. Likewise, Burgess and Barlow (1983) demonstrated that display size and density had little effect on precision in numerosity comparison by a 16-fold change of linear dimensions. As for the effect of element size, Ginsburg and Nicholls (1988), Shuman and Spelke (2006), and Sophian and Chu (2008) demonstrated an inverse relationship between item size and the estimation of numerosity, such as when larger items appeared less numerous than small items. By examining Weber fractions in the numerosity discrimination task, Ross (2003) demonstrated that when the size of elements in comparison was altered, the Weber fractions increased, suggesting that the difference in the size of elements elicited the deterioration of precision in numerosity comparison. Sophian and Chu demonstrated that numerosity judgments were affected by the amount of open space in the array being compared. In addition, recent studies on Stroop interference between discrete (i.e., numerical) and continuous (i.e., element size, cumulative area, or display area) variables have shown inconsistent results (Barth, 2008; Hurewitz, Gelman, & Schnitzer, 2006). Hurewitz et al. argued that interference between discrete and continuous stimulus dimensions occurred in both directions, although it was stronger from the continuous to the discrete; observers respond on the basis of continuous quantity, rather than discrete quantity. On the other hand, Barth provided evidence against Stroop interference between representations of continuous and discrete quantity in judgments of large sets.

Why have so many studies produced inconsistent results? In this study, we attempted to determine the sources of inconsistency and to reconcile the discrepancy in the effects of element size and array area in numerosity judgments. Table 1 presents the experimental conditions and the results of some previous studies. There are at least four factors that could influence the results: task, measurements, number of elements to be compared or estimated, and observers' experience with numerosity judgments. We need to consider how each factor affects the results of experiments.

There are two tasks frequently used for investigating the effects of perceptual variables on numerosity perception: the estimation task and the comparative judgment task. In the estimation task, observers are presented with a stimulus array and are asked to estimate the number of elements in the array; in the comparative judgment task, observers are asked to choose an array consisting of a larger or a smaller number of elements. Estimation and comparative judgment are based on the approximate numerical representation. However, they involve different processes at the response stage: Whereas comparative judgment involves earlier perceptual processes and subsequent decision processes, estimation involves later cognitive processes, such as associating a number with a physical amount (Shepard, 1981). Different processes at the response stage could produce different outcomes for those tasks in numerosity judgments. For example, the recent study by Sophian and Chu (2008) demonstrated that the accuracy of verbal estimates of the numerosities of arrays made upon completion

Table 1
Summary of Experimental Conditions in Previous Studies

Article	Number	Task	Observers (*n*)	Effect on Measurements
Krueger (1972)	25–200	Estimation	Naive (30)	Underestimation of elements in smaller array area
Burgess & Barlow (1983)	25	Comparative judgment	Experienced (2)	No effect of array area
Ginsburg & Nicholls (1988)	19, 37, 61, 117	Estimation	Naive (48)	Overestimation of smaller elements
Vos, van Oeffelen, Tibosch, & Allik (1988)	36	Comparative judgment	Naive (23)	Overestimation of elements in larger array area
Allik, Tuulmets, & Vos, (1991)	16, 32	Comparative judgment	Experienced (2)	No effect of array area
Ross (2003)	8, 16, 32, 64	Comparative judgment	Experienced (5)	Small effect of element size on precision
Shuman & Spelke (2006)	20, 40	Comparative judgment	Naive (24)	Overestimation of smaller elements and elements in larger array area
Sophian & Chu (2008)	40–48	Comparative judgment	Naive (20)	Amount of open space affects the judgment of relative numerosity
	40–48	Estimation	Naive (23)	Amount of open space affects the judgment of relative numerosity

of the comparison task showed little relation to performance on the latter task. Therefore, we proposed consistently using a task (the comparative judgment task) across all the trials to measure the effect of perceptual variables on numerosity judgment exclusively.

As for measurement of the effect of perceptual variables, some studies have measured accuracy (i.e., whether the number of elements was overestimated or underestimated, as compared with the physical amount), some studies have tested precision (i.e., variability of the observer's response or estimation), and a few studies have tested both accuracy and precision. Since perceptual variables could affect both the precision and accuracy of performance, we considered it necessary to measure both precision and accuracy in a particular experiment.

In addition, the number of elements to be compared or estimated is of great importance, for the following reasons. Most investigations of the effects of perceptual variables have used sets of 20 elements or more; smaller sets of elements, such as 5–15, have not been used in many studies. However, the pattern of influence of perceptual variables such as element size and array area differed in the range of numerosity (Durgin, 1995). Since it remains unclear whether the system for representing small numbers of objects is distinct from the system for representing larger numbers of objects, it is necessary to test how and whether the effects of perceptual variables differ among a variety of numerosities.

Another important factor that could affect the results is observers' experience in numerical judgments or related tasks. As is presented in Table 1, the results of experiments in which experienced observers participated have shown little or no effect of element size and array area (Allik et al., 1991; Burgess & Barlow, 1983). On the contrary, studies that have introduced naive or less experienced observers have shown the effects of element size and array area (Ginsburg & Nicholls, 1988; Shuman & Spelke, 2006; Sophian & Chu, 2008). Apart from human studies, it has been reported that extensive training was required, at least in the initial phase of an experiment, in studies that required animals to respond to numerosity under controls for perceptual variables (e.g., Nieder et al., 2002; Tomonaga, 2008). Thus, we predict that experience or practice in numerosity estimation or on the comparison task may influence the effects of element size and array area.

In this study, we examined how perceptual variables affect performance on numerosity judgments under adequate control of the factors mentioned above. The comparative judgment task was used to test the effects of perceptual variables. We introduced a wide range of element numbers, including small and large numerosities, to examine how element number would relate to the effects. We measured both the accuracy and precision of numerosity judgments. We examined whether, under the control of these factors, performance on numerosity judgments would differ between less experienced and experienced observers. Although overviews of previous studies have implied that experienced observers are likely to show no effects of perceptual variables, whereas less experienced

observers will show some effects, few studies have explicitly attempted to examine whether and how practice influences the performance of numerosity judgments in a particular experiment.

In Experiment 1, we introduced the numerosity comparison task to less experienced adult observers in order to examine the effects of element size and array area. We predicted that the effects of perceptual variables would be mostly consistent with the results of previous studies in which less experienced observers had participated. In Experiment 2, we examined whether and how practice would influence precision and accuracy in numerosity comparison. We predicted that precision and accuracy would be enhanced with practice and that the effects of perceptual variables would disappear in the same way as in previous studies in which experienced observers were used. In Experiment 3, we examined whether the effect of practice could transfer to the performance of numerosity discrimination in different stimulus conditions.

EXPERIMENT 1

The purpose of Experiment 1 was to examine the effects of element size and array area on accuracy and precision in numerosity judgments. We employed a wide range of standard element numbers to test whether and how the effects of element size and array area would differ among element numbers. We used the method of constant stimuli, in which observers decided on each trial which visual array, a standard array or a comparison array, had more elements. To test the accuracy of the numerical comparison, we derived the point of subjective equality (PSE). If the size of elements or of the array area affected the numerosity judgment, dissociation between the number of physical elements and perceived numerosity would occur, and the PSE would deviate from the physical amount. To test precision, we derived Weber fractions that indicated the observer's variance of numerosity judgment. Both behavioral and neurobiological evidence have shown that numerosity comparison obeys the Weber law: Discriminability depends on the ratio of the numerosities to be compared (e.g., Brannon, 2006; Burgess & Barlow, 1983; Cantlon et al., 2009; Piazza et al., 2004; van Oeffelen & Vos, 1982). The values of the Weber fraction have been estimated to range from 0.11 to 0.16 when the number of elements tested were in the range of 20–400 in the numerical discrimination of spatial visual arrays (e.g., Allik et al., 1991; Burgess & Balow, 1983; van Oeffelen & Vos, 1982). If the size of the elements or the array area influences the precision of the numerosity comparison, the Weber fraction will deviate from those values.

Method

Participants. Ten observers participated in Experiment 1. All the observers had no prior experience of numerosity comparison. All the observers had normal or corrected-to-normal vision.

Design. Two independent variables were examined in the experiment: perceptual size and the element number of the standard array. Two sets of arrays, a standard array and a comparison array, appeared successively in random order. We used four standard element

numbers: 5, 10, 20, and 40. Because discriminability depends on the ratio of the numerosity to be compared, the ranges of the ratios of the comparison element values to the standard element values were assumed to be identical at all standard element values. We chose six possible comparison values that were symmetric around the standard values so that the ratio of the comparison to the standard values should range from 0.8 to 1.2. The range was chosen so that the data points could cover a wide range on each psychometric function. When the standard element values were 20 and 40, the possible comparison element numbers were 17, 18, 19, 21, 22, and 23 and 34, 36, 38, 42, 44, and 46, respectively. However, when the standard element value was 10, there were only four possible comparison values (8, 9, 11, and 12); when the standard element was 5, there were only two possible comparison values (4 and 6), due to the discrete nature of the dot number. Thus, we chose to use four comparison element values: 8, 9, 11, and 12 at the standard element value of 10 and 3, 4, 6, and 7 at the standard element value of 5, although the range of the ratio was 0.6–1.4.

Two perceptual variables consisted of element size and array area. In the control condition, the sizes of the individual elements and of the display area of the comparison arrays were identical to those in the standard arrays; in the element size condition, the size of the individual elements of the comparison arrays was larger than that of the standard arrays. In the array area condition, the size of the display area of the comparison arrays was larger than that of the standard arrays. Trials in the perceptual variable conditions and the four standard element numbers were intermixed in a block.

Measurements. The PSE and the Weber fractions were measured using the method of constant. First, the number of elements in comparison arrays was plotted on the *x*-axis, and the proportion of "greater" response for each comparison array was plotted on the *y*-axis. Then the plotted data points constructed the psychometric function approximated by a cumulative Gaussian function, on which the difference threshold was obtained. The difference threshold was defined as the smallest amount of the element number change to achieve correct responding of 75%. The Weber fractions were obtained by dividing the difference thresholds by the standard element numbers. The PSEs were obtained as the value of the location on the psychometric function at which the standard and comparative arrays' choice probabilities were equal to 50%. In this experiment, we obtained the standardized PSE, dividing the PSE by the number of standard elements.

Each condition had 480 trials (24 repetitions × 4 levels at standard elements of 5 and 10 and 24 repetitions × 6 levels at standard elements of 20 and 40) and resulted in 1,440 trials in total. Each of the blocks had 120 trials, and there were 12 blocks in total. The sequence of the trials was completely randomized in a block. The standard came first on half of the trials and second on the other half. The observers were given 20 practice trials before the experiment.

Stimuli. The stimuli consisted of light gray dots on a dark gray background. All the dots in a particular array had the same size, but the diameter of the dots in the standard arrays varied, from array to array, between 8 arc min and 12 arc min of visual angle (8, 9, 10, 11, and 12 arc min of visual angle); therefore, the total area of the dots was not a reliable cue to numerosity. The diameter of the dots in the comparison arrays in the control and the array area condition was identical to that of the standard arrays. In the element size condition, the size of the dots in the comparison arrays was 1.5–2.0 (1.5, 1.6, 1.7, 1.8, 1.9, and 2.0) times larger than that of the dots in the standard arrays. The display area of the arrays in the control and the element size conditions subtended 2.8º × 2.8º of visual angle. In the array area condition, the size of the array area of the comparison arrays varied from 1.5 to 2.0 (1.5, 1.6, 1.7, 1.8, 1.9, and 2.0) times larger than that of the standard arrays; therefore, the display area was varied between approximately 3.4º × 3.4º and 4º × 4º of visual angle. We introduced multiple scale factors so that the observers might not notice the exact size difference that could be used to calibrate their responses. The stimuli could be located in any given position in a set of positions with *x*-axis and *y*-axis noise. The sets of the positions also varied from array to array. We controlled the minimum interelement distance[1] and the regularity of spatial distribution[2] of elements, so that the spatial arrangement of the elements was not a reliable cue to numerosity judgments.

Procedure. The observers sat in a darkened room at a distance of approximately 115 cm from the presentation screen. A keypad was placed directly in front of the observers. The observers made responses by pressing the "1" or "3" key. Each trial started with a red fixation cross for 400 msec, followed by the first array. Two sets of arrays, a standard array and a comparison array, were shown in succession in random order. Each array was displayed for 240 msec, separated by an interval for 1,133 msec (a blank for 400 msec, a green square for 333 msec serving as a fixation mark for the second array, and another blank for 400 msec). The observer's task was to answer which array, the first or the second, contained more elements. No feedback about the correctness of the choices was provided. At the beginning of each session, the participants were instructed to judge by the number of elements, and not by other properties of the elements.

A Macintosh G4 computer was used to generate the display and to record the data. The stimuli were presented on a color monitor (SONY Color Graphic Display Model GDM-F400).

Results

The fits of data points to psychometric functions were generally good, and the Pearson product–moment correlation coefficient exceeded .9 in all cases, except for 1 observer. That observer performed two extra blocks, and that correlation coefficient exceeded .9.

Figure 1 shows the Weber fractions and the standardized PSEs of individual observers in each condition. Figure 2 shows the means of the Weber fractions and the means of the standardized PSEs in each condition as functions of standard element numbers. Dotted lines on the figures for the Weber fractions indicate the approximate expected value of the Weber fractions based on previous studies (Burgess & Barlow, 1983; Krueger, 1984; van Oeffelen & Vos, 1982). We termed this value as a criterion Weber fraction. In the present experiment, we assumed the value (i.e., 0.14) as the criterion. Dotted lines on the figures for the PSEs indicated a PSE value of 1.0.

A 3 (condition) × 4 (standard element number) repeated measures ANOVA was conducted on the means of the individual Weber fractions. There were significant main effects of standard element number [$F(3,27) = 17.630$, $p < .01$, $\eta_p^2 = .661$]; a Bonferroni post hoc analysis revealed that the Weber fractions at a standard element number of 5 were significantly different from the others, indicating that the Weber fractions at 5 were substantially smaller than those at other standard numbers ($p < .01$). There was no significant main effect of condition [$F(2,18) = 3.501$, $p > .05$, $\eta_p^2 = .205$], suggesting that precision did not differ across conditions.

In order to test how perceptual variables affect the accuracy of numerosity comparison, a 3 (condition) × 4 (standard element number) repeated measures ANOVA was conducted on the PSEs. There were significant main effects of condition [$F(2,18) = 34.216$, $p < .01$, $\eta_p^2 = .792$]. A Bonferroni post hoc analysis revealed a significant difference across all conditions ($p < .01$), suggesting that the PSEs differed among the control, element size, and
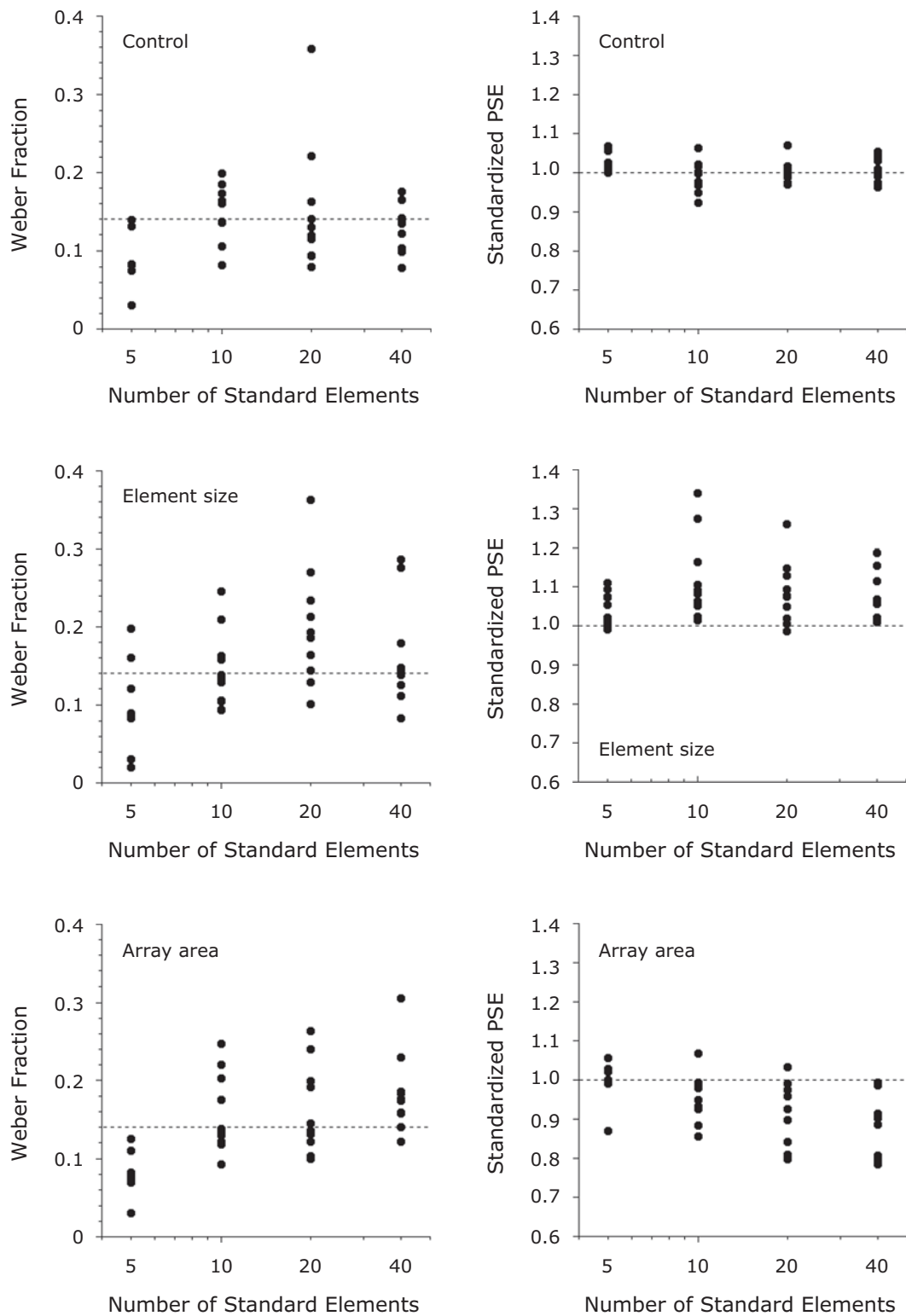
Figure 1. The Weber fractions and the standardized PSEs of individual observers as functions of number of standard element numbers in Experiment 1. "Element size" on the figure indicates element condition: "Array area" indicates array area condition. Dotted lines on the figures of the Weber fractions indicate the approximate expected value of the Weber fractions; dotted lines on the figures of the standardized PSEs indicated the PSE value of 1.0.
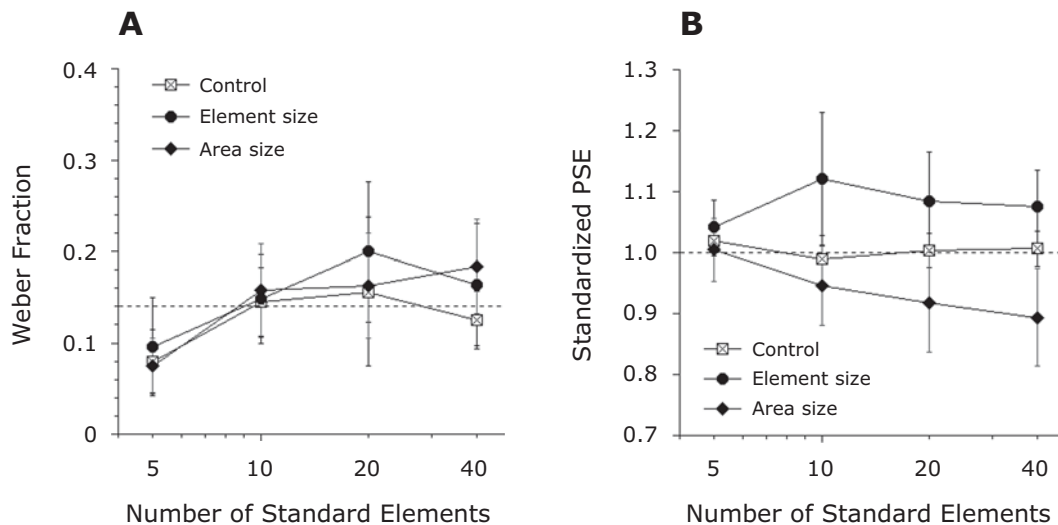
A

B

Figure 2. Mean Weber fraction (A) and standardized PSE (B) as a function of the number of elements in Experiment 1. Error bars represent standard deviations.

array area conditions. A significant interaction between standard element number and condition was obtained as well [$F(6,54) = 8.493, p < .1, \eta_p^2 = .486$], demonstrating that the effects of element size and array area depend on the number of elements.

Furthermore, we carried out a one-sample t test to compare the mean of the PSEs of each standard element number in each condition with the PSE of 1.0. In the control condition, there was no significant difference between the mean of the PSEs and 1.0 at any standard element number. In the element size condition, the mean of the PSE was significantly larger than 1.0 at the standard element numbers of 10 [$t(9) = 3.539, p < .01$], 20 [$t(9) = 3.315, p < .01$], and 40 [$t(9) = 4.018, p < .01$], suggesting an underestimation of larger elements. In the array area condition, the mean of the PSEs was significantly smaller than 1.0 at the standard element numbers of 10 [$t(9) = -2.670, p < .05$], 20 [$t(9) = -3.208, p < .05$], and 40 [$t(9) = -4.224, p < .01$], suggesting an overestimation of elements in the larger array area. In addition, we carried out a repeated measure ANOVA on the variance of the standardized PSEs for each standard element number to demonstrate a significant difference in conditions [$F(2,6) = 9.091, p < .05, \eta_p^2 = .715$]. A Bonferroni post hoc analysis revealed that the variance of the PSEs in the element size ($p < .01$) and area size ($p < .05$) conditions was significantly larger than that in the control condition, suggesting that there was a larger individual difference in accuracy when perceptual variables were manipulated.

Discussion

The results demonstrate that larger elements were judged to be less numerous than smaller elements at the standard element numbers of 10, 20, and 40 and that elements in the larger array area were judged to be more numerous that those in the smaller array area at the standard element numbers of 10, 20, and 40. The results were mostly con-

sistent with those of previous studies in which adult observers with no extensive practice were used (Ginsburg & Nicholls, 1988; Krueger, 1972; Shuman & Spelke, 2006). On the other hand, neither element size nor array area affected the accuracy at the standard number of 5.

In addition, the results demonstrate that precision did not differ across conditions when the standard element number was 10, 20, or 40. However, it should be noted that variability of the Weber fraction for observers was very large; many observers showed values higher than the criterion Weber fraction. One possibility is that some observers did not use all the available numerical information in the arrays. According to Burgess and Barlow (1983), the criterion Weber fraction we presented was derived from the performance of a numerosity discrimination task in which experienced observers made use of the full sample of numerical information available. It could be inferred that the higher Weber fraction obtained in this experiment may have been due to poor use of available numerical information.

More to the point, at the standard element number of 5, the Weber fractions were strikingly smaller than those at the other standard numbers, suggesting that precision was significantly higher than for the other element number conditions.

In the following experiment, we examined whether the effects of perceptual variables found in this experiment would vary with practice.

EXPERIMENT 2

In Experiment 2, we examined whether the effects of perceptual variables found in Experiment 1 would vary with practice. We predicted that practice would enhance precision and reduce bias; the Weber ratio would approximate the expected value in all the conditions, and the PSEs in the element size and array area conditions would converge on the value of 1.0.

Method

Participants. Eight observers who had participated in Experiment 1 had practice sessions and a second test session.

Design. The results of Experiment 1 were considered to be those of the first test session. A practice session consisted of 10 blocks of 120 trials (20 repetitions $\times$ 3 conditions $\times$ 4 levels at standard numbers of 5 and 10, and 20 repetitions $\times$ 3 conditions $\times$ 6 levels at standard numbers of 20 and 40). All the observers had at least three practice sessions. It took about 6 days in total to complete all the sessions. The first practice session was given within 2 days of the first test, which corresponded to Experiment 1. Subsequent practice sessions were given within 2 days. Weber fractions and PSEs were measured at the completion of each session. In cases in which a small amount of improvement in precision and accuracy was observed, extra practice sessions were given. Within 2 days after the practice sessions, the second test (i.e., Test 2) was given. The stimuli and apparatus were the same as those in Experiment 1.

Procedure. The procedure was the same as that in Experiment 1, with the following exception. In the practice session, the observers were given feedback with a short beep sound when they made an incorrect judgment. In the second test, no feedback was provided, as in the first test.

Results

Figure 3 shows the means of the Weber fractions and the means of the PSEs in each condition as a function of standard element number. Solid circles represent the data for Test 1 (i.e., the first test), and blank circles represent the data points for Test 2 (i.e., the second test). The figures show that performance on the second test improved with practice; the Weber fractions approached the criterion value in all the conditions, and the PSEs converged at 1.0 in the element size and array area conditions.

To examine whether practice improved precision in numerosity comparison, a 2 (test) $\times$ 4 (standard element number) $\times$ 3 (condition) repeated measures ANOVA was conducted on the means of the Weber fractions. There was a significant main effect of test [$F(1,7) = 28.535, p < .01$, $\eta_p^2 = .800$], reflecting that practice improved precision in numerosity comparison. Again, there was a significant main effect of standard number [$F(3,21) = 36.192, p < .01, \eta_p^2 = .838$]; a Bonferroni post hoc analysis revealed that there was a significant difference in the Weber fractions between the standard number of 5 and the other standard numbers ($p < .01$). In addition, a 2 (test) $\times$ 4 (standard element number) repeated measures ANOVA was conducted on the variance among the individual Weber fractions in each condition. There was a significant main effect of the test in the element size condition [$F(1,3) = 60.763, p < .01, \eta_p^2 = .783$] and in the array area condition [$F(1,3) = 16.799, p < .05, \eta_p^2 = .463$], suggesting that individual variability in precision was reduced with practice. No significant main effect was observed in the control condition [$F(1,3) = 2.447, p > .05$].

To examine whether practice reduced bias in the element size and array area conditions, we carried out a 2 (test) $\times$ 4 (standard element number) $\times$ 3 (condition) repeated ANOVA on the PSEs. The results indicated significant main effects of condition [$F(2,14) = 20.626, p < .01, \eta_p^2 = .747$] and standard element number [$F(3,21) = 4.147, p < .05, \eta_p^2 = .372$], a two-way interaction between condition and test [$F(2,14) = 9.598, p < .01, \eta_p^2 = .575$], and a three-way interaction between condition, test, and standard element numbers [$F(6,42) = 4.440, p < .01$, $\eta_p^2 = .390$]. The results demonstrated that practice reduced bias in the element size and array area conditions at standard element numbers of 10, 20, and 40.

In addition, we carried out one-sample t tests to compare the means of the PSEs for each standard element number in each condition with the PSE of 1.0 in Test 2 in order to examine whether the bias was eliminated. In the element size condition, the bias observed in Experiment 1 was eliminated at the standard numbers of 10 [$t(7) = 2.330, p > .05$], 20 [$t(7) = 0.973, p > .05$], and 40 [$t(7) = 1.171, p > .05$], suggesting no underestimation of the number of larger elements. In the array area condition, the bias observed in Experiment 1 was eliminated at the standard numbers of 10 [$t(7) = -1.837, p > .05$] and 40 [$t(7) = -1.416, p > .05$] as well, suggesting no overestimation of the number of elements in the larger array area. At the standard number of 20, overestimation of the larger array area remained [$t(7) = -2.501, p < .05$], even though there was a reduction in the extent of overestimation.

Furthermore, we examined how practice influenced accuracy and precision in numerosity comparison by tracing the individual processes of 2 observers. Figure 4 shows the practice effect of 1 of the observers whose performance showed the typical practice effect, characterized by a decrease of Weber fractions values and the elimination of bias. The Weber fractions approached the level of the criterion value for all the standard numbers, demonstrating that precision was enhanced with practice. Similarly, the PSEs in both the element size and array area conditions converged on the value of 1.0, demonstrating the reduction of bias with practice. Figure 5 shows the process of 1 observer who demonstrated a small improvement in performance. The observer was given an extra practice session. Although the Weber fractions gradually decreased with the observer's practice in all the conditions, they did not reach the expected value. The PSEs in the array area condition persistently remained at the same level even with extra practice by the observer. In Test 2, only a slight improvement in precision was observed, whereas the PSEs remained almost at the same level as in Test 1.

Discussion

The results demonstrate that precision was improved in all the conditions with practice and that the effects of element size and array area found in Test 1 disappeared in Test 2. The findings provide clear evidence that practice influenced accuracy and precision in the numerosity comparison. Thus, observers could, with practice, discriminate between two visual arrays on the basis of numerosity. Consequently, we can infer that the discrepant results for perceptual variables in previous studies may have been due to differences in practice, at least in part.

Two significant questions arise. First, when we examined the practice effects across sessions individually, 2 out of 8 observers showed persistent bias in the array area condition despite extended practice. Why did 2 observers, both of whom showed numerical overestimation of elements in larger array areas, persistently show array area bias? A possible explanation is that the materials and
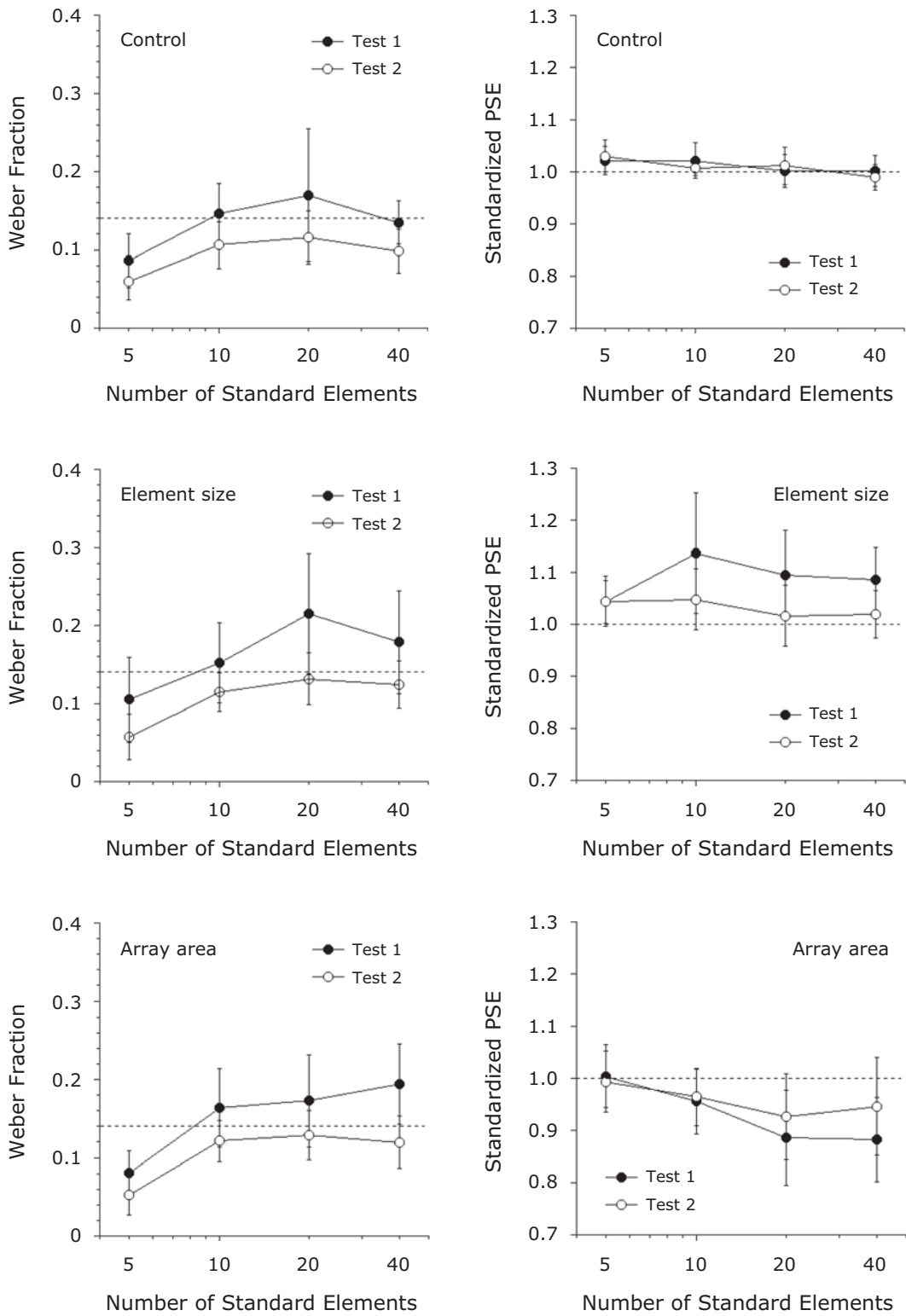
Figure 3. Mean Weber fraction (left panels) and standardized PSE (right panels) as a function of the number of elements in Test 1 and Test 2 in Experiment 2. Error bars represent standard deviations.
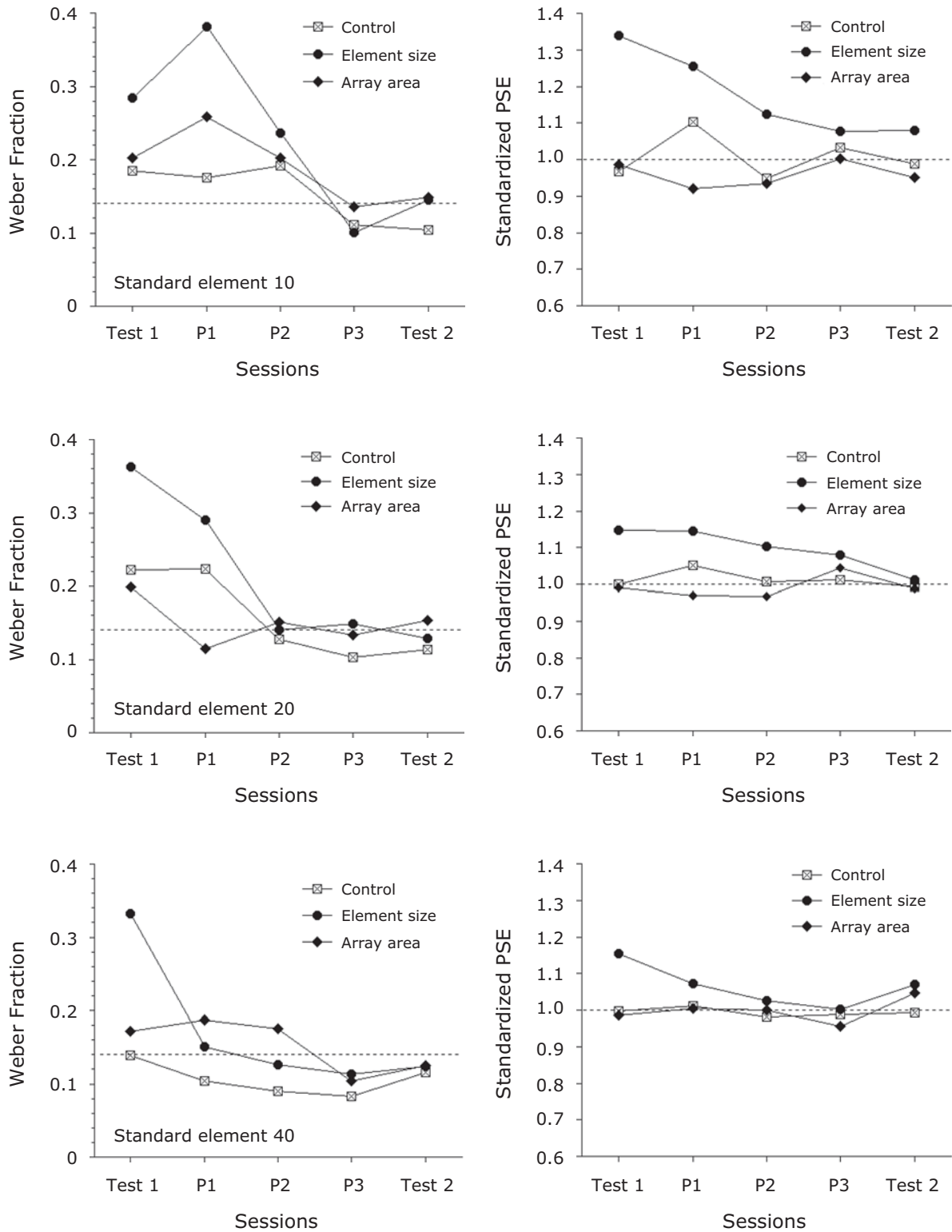
Figure 4. The transitions of the Weber fractions and the standardized PSEs of Observer B as functions of tests and practice sessions in Experiment 2. P1, P2, and P3 indicated the first, second, and third practice sessions.
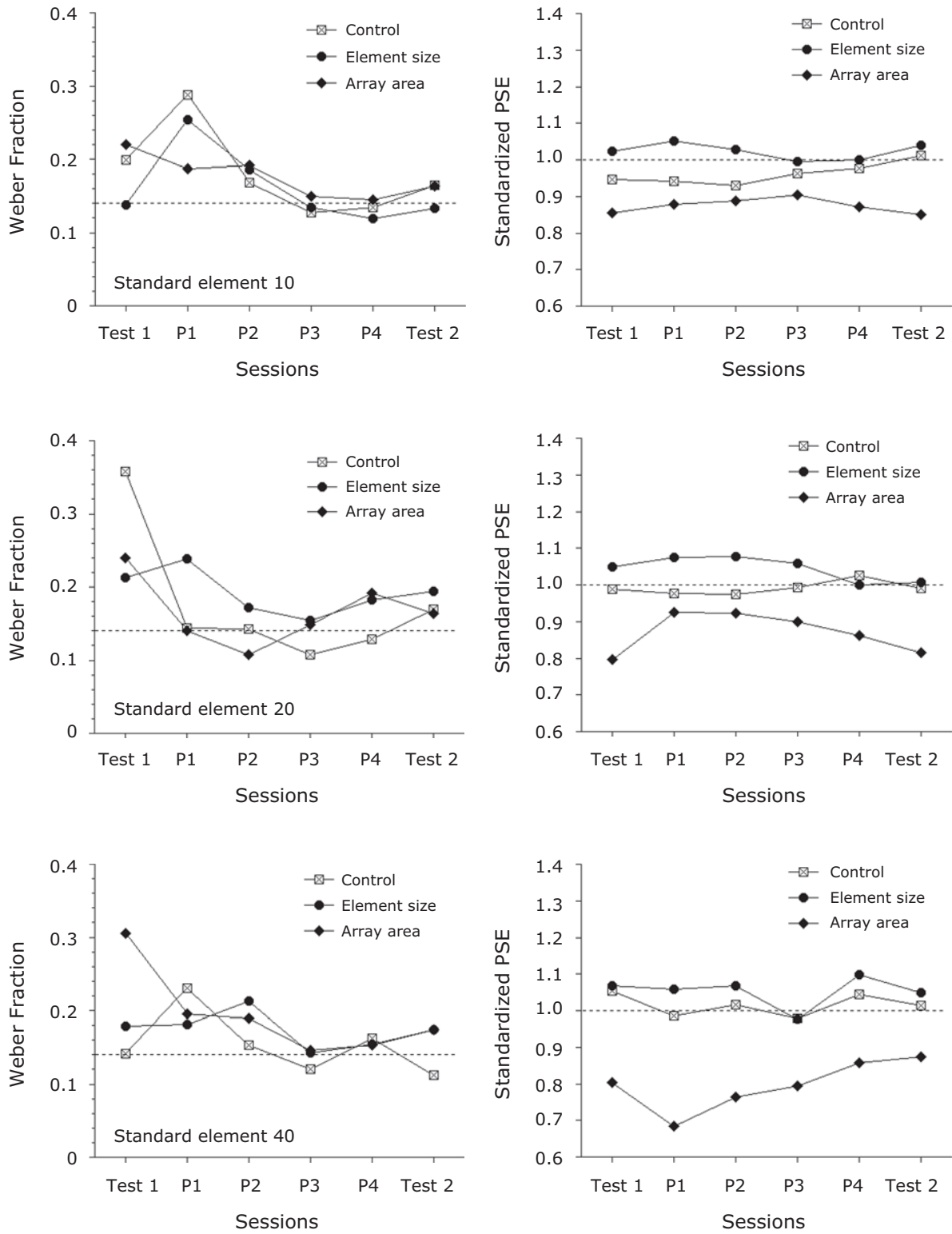
Figure 5. The transitions of the Weber fractions and the standardized PSEs of Observer C as functions of tests and practice sessions in Experiment 2. The observer served in the extra practice session (P4).

procedures in the practice sessions were not appropriate for guiding them to attend exclusively to the numerical dimension. Although feedback had shown the observers' judgments to be incorrect on each trial, that feedback did not indicate how they were wrong. Since we used identical numerical contrasts of standard and comparison arrays across all trials, the difficulty of the task in the practice sessions was the same as that on the test trials. For some of the observers, the difficulties of the trials should be modified on the practice trials, so that they can know "what it is like to attend to the numerosity" systematically. In further investigations, we need to design a more appropriate practice procedure to test whether all observers can eliminate bias and improve precision. Another possibility is that the observers may have failed to learn to make numerosity discrimination because they did trials quickly and carelessly and did not pay much attention to whether they were right or wrong. To test the possibility, we need to obtain response time data in further investigations.

Another important question is what, exactly, the observers learned through the practice. One possibility is that they learned to abstract truly numerical information by selectively paying attention to the numerosity of elements by inhibiting the interference from the magnitude of irrelevant dimensions, such as total surface area or display area. If observers have been trained to extract truly numerical values, they should show the same performance level in a task in which different stimulus conditions are used. In the following experiment, we examined whether the effect of practice would transfer to different stimulus conditions in which the size of the element and the array area in comparison arrays varied by different scale factors.

EXPERIMENT 3

The purpose of this experiment was to examine whether the practice effect found in Experiment 2 would transfer to other perceptual values of stimuli. We introduced the *same* and *transfer* conditions. In the *same* condition, the same sets of elements appeared both in the practice session and in the test session; in the transfer condition, different sets of elements appeared in the two sessions. The stimuli for the test session in the *same* condition and those in the transfer condition differed with respect to the scale factors.

If observers learned to extract truly numerical values with practice, they would show the same performance level even in the task using stimuli with different scale factors. On the other hand, if observers learned to perform a task with stimuli with a specific scale factor, their performance in the task using stimuli with different scale factors would not be improved with practice.

Method

Participants. Eight observers participated in this experiment. All the observers had normal or corrected-to-normal vision.

Design. In this experiment, the standard element numbers of 20 and 40 were tested, since no bias was observed with the standard element number of 5 in any condition and the trend for bias with the standard element number of 10 was similar to that for 20. The element size and array area conditions were examined.

The experiment consisted of five sessions: the first test session, three practice sessions, and the second test session. Weber fractions and PSEs were analyzed as a function of the size of the standard element number (20, 40), stimulus condition (same, transfer conditions), and perceptual variables (element size, area size). The first test session was composed of 10 blocks of 96 trials (2 repetitions × 6 levels × 2 standard numbers × 2 perceptual variables × 2 stimulus conditions). A practice session was composed of 4 blocks of 96 trials (4 repetitions × 6 levels × 2 standard numbers × 2 perceptual variables). All observers had three practice sessions in total. The second test session was composed of 10 blocks of 96 trials (2 repetitions × 6 levels × 2 standard numbers × 2 perceptual variables × 2 stimulus conditions), as in the first test session.

Stimuli. The stimuli and apparatus were the same as those in Experiments 1 and 2, with the following exceptions. We explained the stimuli in the first and second test sessions. In the *same* condition, the diameter of the dots in the comparison arrays in the element size condition was 2.0 times larger than that in the standard arrays; the array area size of the comparison arrays in the array area condition was 2.0 times larger than that of the standard arrays. In the transfer condition, the diameter of the dots in the comparison arrays in the element size condition was 4.0 times larger than that in the standard arrays; the array area size of the comparison array in the array area condition was 4.0 times larger than that of the standard arrays. For the practice sessions, we used the same stimuli as those in the test session in the *same* condition.

Procedure. The procedure was the same as that in Experiments 1 and 2, with the following exception. In the test sessions, no feedback was provided. In the practice sessions, the observers were given feedback with a short beep sound when they made an incorrect judgment.

Results

Figures 6 and 7 show the means of the Weber fractions and the means of the standardized PSEs of 8 observers, respectively. Solid circles and diamonds with solid lines show the results in the *same* condition (i.e., scale factor of 2.0). Blank circles and diamonds with dotted lines show the results in the transfer condition (i.e., scale factor of 4.0).

Weber fraction. In the *same* condition, the Weber fractions in the element size and array area conditions for all the observers decreased with practice and approached the criterion value, suggesting that precision in numerosity discrimination improved with practice. Similar transitions were observed in the transfer condition. To examine whether the practice effect would transfer across scaling factors, we conducted a 2 (standard element number) × 2 (scale factor) × 2 (test) repeated measures ANOVA in the element size condition and array area condition separately. The analysis resulted in a significant test effect in both the element size [$F(1,7) = 28.770$, $p < .01$, $\eta_p^2 = .807$] and the array area [$F(1,7) = 38.332$, $p < .01$, $\eta_p^2 = .845$] conditions. In the transfer condition, the Weber fractions in the element size and array area conditions decreased and approached the criterion value in a similar manner as in the *same* condition. The results demonstrate that the practice effect transferred to the stimuli with different scale factors, both in the element size condition and in the array area condition.

PSE. In the *same* condition, the PSEs converged at a value of 1.0 in both the element size and array area conditions, suggesting that accuracy in numerosity discrimination improved with practice. Similar transitions were
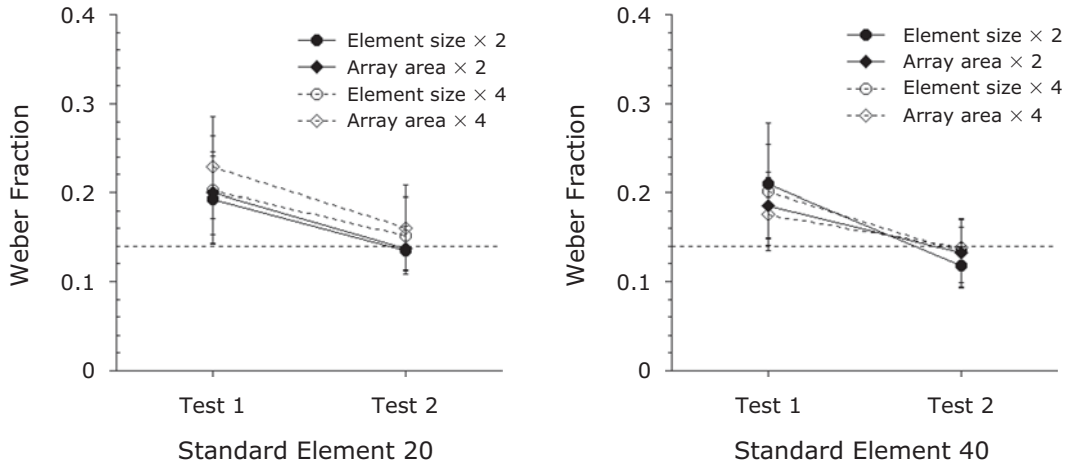
Figure 6. Mean Weber fraction in each condition in the *same* condition (i.e., element size × 2 and array area × 2) and the transfer condition (i.e., element size × 4 and array area × 4) in Experiment 3. Dotted lines on the figures indicate the approximate expected value of the Weber fraction.
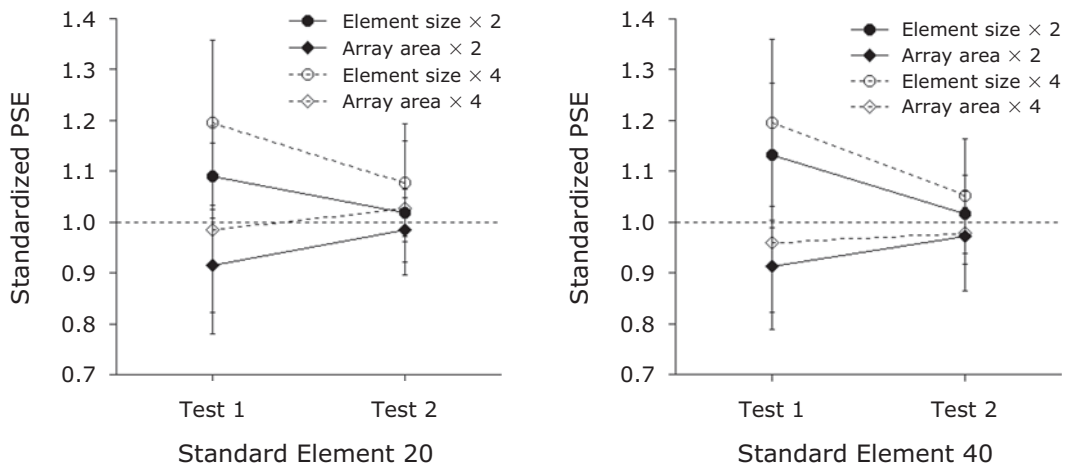

Figure 7. Mean standardized PSE in the *same* condition (i.e., element size × 2 and array area × 2) and the transfer condition (i.e., element size × 4 and array area × 4) in Experiment 3. Error bars represent standard deviations. Dotted lines on the figures indicate the standardized PSE value of 1.0.

observed in the transfer condition. To examine whether the bias pattern shown in Experiment 1 was also observed in each perceptual variable condition, we carried out a one-sample t test to compare the mean of the standardized PSEs at standard element numbers of 20 and 40 with 1.0. In the element size condition, the mean of the PSE was significantly larger than 1.0 for the standard element numbers of 20 [$t(7) = 3.958$, $p < .01$] and 40 [$t(7) = 2.628$, $p < .05$] in the *same* condition and for 20 [$t(7) = 3.421$, $p < .05$] and 40 [$t(7) = 3.349$, $p < .05$] in the transfer condition, suggesting an underestimation of larger elements. The results were consistent with the bias found in Experiment 1. In the array area condition, the means of the PSE was significantly smaller than 1.0 for the standard element numbers of 20 [$t(7) = -2.589$, $p < .05$] and 40 [$t(7) = -2.708$, $p > .05$] in the *same* condition, suggesting an underestimation of larger array area. The results were consistent with the bias found in Experiment 1. In

the *transfer* condition, however, the means of the PSE was not significantly different from 1.0 for the standard element numbers of 20 [$t(7) = -0.221$, $p > .05$] and 40 [$t(7) = -0.672$, $p > .05$], which were not consistent with the bias found in the *same* condition.

To examine whether the practice effect would transfer across scaling factors, we conducted a 2 (standard element number) × 2 (scale factor) × 2 (test) repeated measures ANOVA on PSEs in the element size and array area conditions, respectively. The analysis resulted in a significant test effect [$F(1,7) = 12.524$, $p < .01$, $\eta_{p}^{2} = .641$] in the element size condition, suggesting that the practice effect transferred across scaling factors. However, no significant test effect was observed in the array area condition [$F(1,7) = 1.508$, $p > .1$]: A test effect was observed only in the *same* condition [$F(1,7) = 5.728$, $p < .05$, $\eta_{p}^{2} = .479$]. Since the overestimation of element numbers in the larger transfer condition (i.e., scale factor of 4) was not

observed, we could not have an expectation of whether the reduction of bias in the *same* condition would transfer across scaling factors. We carried out a one-sample t test to compare the means of the PSEs for test in the *same* and transfer condition with 1.0 to demonstrate no bias for Test 2 in the array area condition for standard element numbers of 20 [$t(7) = -0.695, p > .1$] and 40 [$t(7) = -1.362, p > .1$] in the *same* condition and for standard element numbers of 20 [$t(7) = 0.595, p > .1$] and 40 [$t(7) = -0.518, p > .1$] in the transfer condition.

Discussion

The results demonstrate that the Weber fractions in the transfer condition decreased to the level of those in the *same* condition. Namely, the observers showed the same performance level even in the task using stimuli with different scale factors. The results also demonstrate that the standardized PSEs converged on a value of 1.0 and all bias was eliminated after practice sessions. Thus, we could predict that the observers learned to extract numerical values with practice.

It should be noted that the direction of bias in the transfer condition (i.e., scale factor of 4) in the array area condition was not consistent with the findings in the *same* condition. One possible explanation is that the size of the scale factors may influence the size of bias. Elements in a much larger area might give the impression of "sparse" distribution for some observers. Further investigation is needed to explain the discrepancy.

GENERAL DISCUSSION

We attempted to determine the source of the inconclusive evidence in previous studies about the effects of element size and array area on numerosity judgments. We demonstrated that precision and accuracy improved with practice and that the effects of these perceptual variables could be eliminated.

In Experiment 1, we tested the effects of element size and array area in a numerosity comparison by use of a wide range of numbers of elements. The results showed that element size and array area affected precision and accuracy in a numerosity comparison. Naive observers showed Weber fractions larger than the criterion value, reflecting lower precision in the numerosity comparison. As for accuracy, the larger elements were judged to be less numerous than the smaller elements at the standard element numbers of 10, 20, and 40; elements in the larger array area were judged to be more numerous than those in the smaller array area at the standard numbers of 10, 20, and 40; neither effect was observed with the standard number of 5. The results are mostly consistent with those of previous studies that used observers with no extensive practice, shown in Table 1 (Ginsburg & Nicholls, 1988; Krueger, 1972, 1984; Shuman & Spelke, 2006). It should be noted that when the number of elements was about 5 or smaller, no effect of element size and array area on precision and accuracy was observed, irrespective of practice. There are two possibilities for explaining the difference in data patterns at the standard element number of 5. One

is that the smaller and larger numerosities mediate one continuous form of representation and the perceptual variables have a graded effect on numerical judgments, with an increasing impact as the number increases. Another possibility is that the processing of a smaller numerosity and a larger numerosity mediates independent systems (e.g., Peterson & Simon, 2000; Trick & Pylyshyn, 1994). However, the possibility is less likely, because we used comparison values of 3, 4, 6, and 7 to obtain the Weber fraction and PSE at the standard value of 5, which are mixtures of values within and outside the subitizing range.

In Experiment 2, we examined whether practice would influence accuracy and precision in a numerosity comparison. Our results showed that precision improved and that bias was reduced with practice in most cases. Specifically, the Weber fractions decreased to the level of the criterion value, and the PSEs converged on a value of 1.0. The findings suggested that the observers learned to discriminate the arrays on the basis of the numerical value and were unaffected by the perceptual variables. The results were consistent with those in the previous studies that involved observers with extensive practice (Allik et al., 1991; Burgess & Barlow, 1983).

In Experiment 3, we examined whether the practice effect found in Experiment 2 would transfer to other perceptual values of stimuli. It has been argued that transfer of the practice effect implies that related tasks are learned on a relatively higher level of information processing, whereas the specificity of the practice effect implies involvement of lower level processing. Since the observers showed the same performance level in the task using stimuli with different scale factors, we inferred that the observers had learned to extract numerical values with practice.

Taken together, our findings provide clear evidence that the inconsistent effects of perceptual variables found in previous studies were due to the difference in the observers' experience in numerical tasks. More specifically, naive observers tend to show a larger bias and lower precision; experienced observers tend to show little bias and higher precision.

Two significant questions arise. First, how are naive observers affected by perceptual variables in their numerosity judgments? Why are naive observers susceptible to the effect of perceptual variables even though they are equipped with a numerical representation system that could extract a true numerical value? The fact that the observers in Experiment 1 demonstrated a high variability of precision and accuracy in performance across conditions suggests that the observers did not make a numerosity comparison solely on the basis of the neural mechanisms for the numerical representation but, rather, made it by incorporating perceptual factors such as display size and element size. One possibility is that although the neuronal mechanisms for generating approximate numerical representation could be used in numerosity comparison, the representations of other perceptual magnitude, such as spacing, area size, and/or element sizes, may interfere when a decision is made (Hurewitz et al., 2006). For example, sets of elements arranged in a larger display area may be judged to be more numerous because the magnitude of a display area

interferes with the numerosity of the elements. However, the fact that observers are likely to judge smaller elements as being more numerous than larger ones contradicts the idea that the magnitude of an aggregated area interferes with numerosity judgments, since observers should judge larger elements to be more numerous when the magnitude interferes. The result may support the claims made by Barth (2008), providing evidence against stroop interference of continuous values with numerical values in terms of element size and the aggregated area. One of the other continuous variables, spacing, defined as the average distance between the circumferences of an element and that of the nearest element, might have interfered with numerosity. Size of spacing is inversely correlated with change of element size when the minimum interelement distance and the array area are controlled (i.e., larger elements produce smaller spacing). In this case, the set of elements with larger spacing (i.e., smaller elements) could be judged to be more numerous than that with the smaller spacing (i.e., larger elements). To test this possibility, the systematic control of spacing and the size of elements will be required in further investigations.

An alternative possibility is that observers can be influenced by cognitive or decision processes. Learned knowledge, such as "a larger area is supposed to have a large number of elements" or "smaller elements are usually more numerous than larger ones," may influence numerosity judgments. Sophian and Chu (2008) pointed out that the knowledge that a smaller size is often associated with a larger numerosity in daily experience led observers to judge an array composed of smaller elements as more numerous.

Probably, the interference of the magnitude of other perceptual properties and/or the influence of learned knowledge on the decision process could be the source of the effects of perceptual variables for naive observers. A numerosity comparison involves multiple processing stages, including the earlier perceptual and the later decision-making processes. Bias and a decline in precision may occur at any of those stages, which depend on the observers' experience and/or knowledge of numerical judgments.

Second, how are the effects of perceptual variables eliminated with practice? The fact that the practice effect on particular stimuli transferred to different stimuli implies that the participants seemed to learn to abstract a numerical value. A plausible explanation is that observers could learn to suppress the interference of other perceptual variables and/or the influence of prior knowledge on numerical judgment with practice. Consequently, they could attend to the numerical properties of spatial arrays exclusively in order to abstract the true numerical values.

For further investigation, we need to examine the effect of perceptual variables other than element size and display area. Approximate numerical representation may be abstract, but the process of extracting numerical information requires perceptual information, such as density of elements, cumulative area, spacing, and so forth. Many studies have pointed out that some properties of spatial arrangements of element sets affect a numerosity judgment (e.g., Allik & Tuulmets, 1991; Ginsburg, 1991; Sophian &

Chu, 2008). For example, regularly distributed elements in an array are judged to be more numerous than irregularly distributed elements in an array (Allik & Tuulmets, 1991; Ginsburg, 1991). We need to test whether the effect of spatial arrangement is eliminated with practice. Furthermore, we need to test whether the practice effect on element size and array area could transfer to stimuli in a different spatial arrangement.

In conclusion, we provide clear evidence that difference in the observers' experiences is the source of the discrepant effects of element size and array area. The finding that precision and accuracy improved with extensive practice suggest that extracting the true numerical value is possible only through practice and/or experience, even for human adults. In order to provide a clear description of an approximate numerical representation system, we need to examine how our numerical judgment is affected by perceptual variables and how, with practice, we learn to extract a true numerical value.

AUTHOR NOTE

REFERENCES

Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics*, **49**, 303-314.

Allik, J., Tuulmets, T., & Vos, P. G. (1991). Size invariance in visual number discrimination. *Psychological Research*, **53**, 290-295.

Barth, H. C. (2008). Judgments of discrete and continuous quantity: An illusory Stroop effect. *Cognition*, **109**, 251-266. doi:10.1016/j.cognition.2008.09.002

Brannon, E. M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology*, **16**, 222-229.

Burgess, A., & Barlow, B. H. (1983). The precision of numerosity discrimination in arrays of random dots. *Vision Research*, **23**, 811-820.

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, **4**, e125.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, **13**, 83-91. doi:10.1016/j.tics.2008.11.007

Cappelletti, M., Barth, H., Fregni, F., Spelke, E. S., & Pascual-Leone, A. (2007). rTMS over the intraparietal sulcus disrupts numerosity processing. *Experimental Brain Research*, **179**, 631-642.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science*, **10**, 408-411.

Cordes, S., & Brannon, E. M. (2008). The difficulties of representing continuous extent in infancy: Using number is just easier. *Child Development*, **79**, 476-489.

Cordes, S., & Brannon, E. M. (2009). The relative salience of discrete and continuous quantity in young infants. *Developmental Science*, **12**, 453-463. doi:10.1111/j.1467-7687.2008.00781.x

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, **21**, 355-361. doi:10.1016/S0166-2236(98)01263-6

Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 149-169. doi:10.1037/0096-1523.21.1.149

EGER, E., STERZER, P., RUSS, M. O., GIRAUD, A. L., & KLEINSCHMIDT, A. (2003). Supramodal number representation in human intraparietal cortex. *Neuron*, **37**, 719-725.

FEIGENSON, L., DEHAENE, S., & SPELKE, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, **8**, 307-314. doi:10.1016/j.tics.2004.05.002

GINSBURG, N. (1991). Numerosity estimation as a function of stimulus organization. *Perception*, **20**, 681-686.

GINSBURG, N., & NICHOLLS, A. (1988). Perceived numerosity as a function of item size. *Perceptual & Motor Skills*, **67**, 656-658.

HOLLINGSWORTH, W. H., SIMMONS, J. P., COATES, T. R., & CROSS, H. A. (1991). Perceived numerosity as a function of array number, speed of array development, and density of array items. *Bulletin of the Psychonomic Society*, **29**, 448-450.

HUREWITZ, F., GELMAN, R., & SCHNITZER, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences*, **103**, 19599-19604.

KRUEGER, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, **11**, 5-9.

KRUEGER, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & Psychophysics*, **35**, 536-542.

MIX, K. S., HUTTENLOCHER, J., & LEVINE, S. C. (2002). Multiple cues for quantification in infancy: Is number one of them? *Psychological Bulletin*, **128**, 278-294.

NIEDER, A., & DEHAENE, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, **32**, 185-208.

NIEDER, A., FREEDMAN, D. J., & MILLER, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, **297**, 1708-1711.

NIEDER, A., & MERTEN, K. (2007). A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *Journal of Neuroscience*, **27**, 5986-5993.

NIEDER, A., & MILLER, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, **37**, 149-157. doi:10.1016/S0896-6273(02)01144-3

PETERSON, S. A., & SIMON, T. J. (2000). Computational evidence for the subitizing phenomenon as an emergent property of the human cognitive architecture. *Cognitive Science*, **24**, 93-122. doi:10.1016/S0364-0213(99)00022-1

PIAZZA, M., IZARD, V., PINEL, P., LE BIHAN, D., & DEHAENE, S. (2004). Turning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, **44**, 547-555.

ROSS, J. (2003). Visual discrimination of number without counting. *Perception*, **32**, 867-870.

ROUSSELLE, L., PALMERS, E., & NOEL, M. P. (2004). Magnitude comparison in preschoolers: What counts? Influence of perceptual variables. *Journal of Experimental Child Psychology*, **87**, 57-84.

SHEPARD, R. N. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology*, **24**, 21-57.

SHUMAN, M., & SPELKE, E. (2006). Area and element size bias numerosity perception. *Journal of Vision*, **6**(6), 777a.

SOPHIAN, C., & CHU, Y. (2008). How do people apprehend large numerosities? *Cognition*, **107**, 460-478. doi:10.1016/j.cognition.2007.10.009

TOMONAGA, M. (2008). Relative numerosity discrimination by chimpanzees (*Pan troglodytes*): Evidence for approximate numerical representations. *Animal Cognition*, **11**, 43-57.

TRICK, L. M., & PYLYSHYN, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, **101**, 80-102.

VAN OEFFELEN, M. P., & VOS, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics*, **32**, 163-170.

VOS, V. G., VAN OEFFELEN, M. P., TIBOSCH, H. J., & ALLIK, J. (1988). Interactions between area and numerosity. *Psychological Research*, **50**, 148-154. doi:10.1007/BF00310175

WHALEN, J., GALLISTEL, C. R., & GELMAN, R. (1999). Non-verbal counting in humans: The psychophysics of number representation. *Psychological Science*, **10**, 130-137.

NOTES

1. In this study, interelement distance is defined by the distance from a center of an element to those of other elements. For the interelement distance controls, we set the minimum interelement distance (MinIE). We had two levels of MinIE for the standard element number 5 and three levels of MinIE for the standard numbers 10, 20, and 40, so that the spatial configuration would not be the cue for numerosity judgments.

2. To ensure that the influence of element distribution did not differ across display sizes, we controlled the relative regularity of element distribution by introducing the ratio of relative regularity. The ratio of relative regularity is expressed by MinIE/MinIEmax (Durgin, 1995), where MinIEmax is the theoretical maximum MinIE expressed as a square root of area/number of elements.