

Unstable mean context causes sensitivity loss and biased estimation of variability

State Key Laboratory of Brain and Cognitive Science,
Institute of Psychology, Chinese Academy of Sciences,
Beijing, China
University of Chinese Academy of Sciences,
Beijing, China

Ke Tong



Psychopathology and Affective Neuroscience Laboratory,
Department of Experimental-Clinical and Health Psychology,
Ghent University, Ghent, Belgium

Luyan Ji



State Key Laboratory of Brain and Cognitive Science,
Institute of Psychology, Chinese Academy of Sciences,
Beijing, China

Wenfeng Chen



State Key Laboratory of Brain and Cognitive Science,
Institute of Psychology, Chinese Academy of Sciences,
Beijing, China

Xiaolan Fu



A recent study has suggested that statistical representations of ensemble objects may provide contextual stability to facilitate perception. The present study investigated whether facilitating such perception occurs in the extraction of variability information and how the stability of context mean values influences variability perception. We designed two tasks in which participants directly judged the variability of stimuli. In Experiment 1, we manipulated both the stability of the mean values and the exposure time to observe the time course of stability facilitation. In Experiment 2, we decomposed the stability of the context mean values into between-trials and within-trial levels to further investigate the mechanism of such facilitation. The results revealed that stable mean contexts do facilitate variability perception. In particular, unstable long-term mean context causes loss of sensitivity to variability whereas response bias is determined by the interaction between long-term and transient mean stability.

effectively capture the degree of central tendency and dispersion. Similarly, our perception system is able to extract central tendency and variability information from an overwhelming amount of sensory input to form ensemble statistics (for a review, see Alvarez, 2011). This capability has been tested using different types of stimuli in multiple sensory modalities (recent examples: Schweickert, Han, Yamaguchi, & Fortin, 2014; Sweeny & Whitney, 2014), and the results suggest that it may be a common mechanism in human perception.

Variability: The understudied ensemble statistics

Variability perception is vitally important in terms of survival. For example, camouflage, which is highly survival-related in both natural and artificial environments, relies on the basic principle of manipulating irrelevant variation to mask the detection of relevant variation (Morgan, Mareschal, Chubb, & Solomon, 2012). Consequently, understanding variability perception—and especially understanding the encoding mechanisms behind it—is the key to deciphering the hidden information under camouflage. Variability

Introduction

We frequently use the concepts of mean and standard deviation (*SD*) to describe data sets because they

Citation: Tong, K., Ji, L., Chen, W., & Fu, X. (2015). Unstable mean context causes sensitivity loss and biased estimation of variability. *Journal of Vision*, 15(4):15, 1–12, doi:10.1167/15.4.15.

perception is also important in more general decision-making scenarios. Because the central tendency features (such as mean value and mode) do not cover the degree of dispersion in the data set, tasks that require evaluation of dispersion—such as comparing the homogeneity of different sets—depend on the perception of variability (Tong, Tang, Chen, & Fu, 2015).

However, most current studies of ensemble statistics have focused on the representation of the average properties of the stimulus set whereas perception of variability has been less investigated. Recent studies have shown that matching the variability of the prime and the target might facilitate the processing of the target even if variability is task-irrelevant, which suggests that humans might automatically extract variability information from a stimulus set (Michael, de Gardelle, & Summerfield, 2014). The variety of the stimulus types used in variability studies has increased (e.g., texture in Dakin, Mareschal, & Bex, 2005; biological motion in Sweeny, Haroz, & Whitney, 2013), but the volume of variability studies is dwarfed by the number of active investigations of mean representations.

Most previous studies regarding variability perception have adopted tasks that do not require participants to directly respond to variability information. Thus, the characteristics of variability perception must be inferred from its effects on performance that are based on other features (Marchant, Simons, & de Fockert, 2013; Michael et al., 2014). To better understand variability processing, we must design and perform experiments that require participants to directly access variability information.

Facilitation effect of statistical stability

Numerous studies have investigated the contents of and mechanisms involved in ensemble statistics; however, less is known about their functional roles. One recent study has suggested that ensemble statistics may serve to build perceptual stability and facilitate ongoing perception (Corbett & Melcher, 2014b). Because of our limited visual attention, we only process a small portion of a visual scene in detail, and the majority remains a rough image. However, the statistical characteristics of this coarse image are crucial for us to recognize the background as a frame of reference to maintain visual stability when we switch our focus. For example, both target discrimination speed and scene-scanning fluency were enhanced in contexts with stable global mean values (Corbett & Melcher, 2014b). Within the facilitation effect, ensemble statistics may be extracted using nonselective visual processing and may serve to guide selective visual processing, such as object recognition and visual search (Wolfe, Võ, Evans, & Greene, 2011). It has been suggested that such guidance may function by

predicting the target location based on scene context and by freeing attentional resources to more effectively find targets (Corbett & Melcher, 2014b).

Aims of the present study

Our growing understanding of the processing mechanism behind ensemble statistics derives primarily from mean representation studies, such as parallel versus sampling dialogue (Ariely, 2008; Chong, Joo, Emmmanouil, & Treisman, 2008; Myczek & Simons, 2008). In contrast, little is known about the mechanism of variability processing and its relationship with mean representation. Two important questions in this regard concern how humans encode variability and whether variability and mean information are processed via shared or separate mechanisms. The “priming by variance” phenomenon (Michael et al., 2014) reveals the impact of variability processing on mean representation and may be one consequence of a shared mechanism of mean and variability processing. If so, we might expect that processing mean representation would also influence variability perception.

To address these issues, we aimed to directly evaluate participants’ responses to variability information and determine how variability processing interacts with mean processing. Corbett and Melcher (2014b) led us to investigate this issue via the contextual stability facilitation effect; in their study, backgrounds with stable mean values (a global feature) facilitated the visual search for a singleton among the tilted Gabor patches (a local feature). Because the stability of mean values might facilitate the processing of a local feature, it naturally follows to ask whether the stability of mean values might also facilitate the processing of another global feature, i.e., variability. Answering this question is important because little is known about the interplay between different global characteristics.

In our study, two experiments manipulated the stability of contextual mean values in variability comparison tasks using simultaneous and successive display paradigms. In Experiment 1, the stimulus exposure time was manipulated to determine the time course of the stability facilitation effect. In Experiment 2, the context mean stability was further divided into between-trials and within-trial levels to identify its effect on variability perception.

Experiment 1

In Experiment 1, the participants were asked to compare the variability of two images that were presented simultaneously. The stability of the mean

values was manipulated to investigate the stability facilitation effect. The exposure time of the images was also manipulated to obtain both the time course of variability perception and the stability facilitation effect.

Our first hypothesis was that human participants might quickly extract variability information (Michael et al., 2014), which would be reflected in an above-chance accuracy within a brief exposure duration. The second hypothesis was that a stable context of mean values would provide a processing advantage in the variability judgment task. Dissociation of performance in the stable and unstable blocks may be observed based on the proposed stability facilitation.

Methods

Participants

A total of 15 naïve participants (nine females ranging in age from 20 to 26 years with a mean of 23.4 years and *SD* of 1.9 years) were recruited from nearby universities. All the participants had normal or corrected-to-normal vision. All the participants received monetary rewards for completing the experiment.

Apparatus

The stimulus was presented using a Philips 109B6 17-in. CRT display (Philips, Amsterdam, Netherlands) with a resolution of 1024×768 and a 100-Hz refresh rate. The experimental procedure was controlled using the E-Prime 2.0 software package (Psychology Software Tools, Pittsburgh, PA), and the responses were collected using a standard keyboard. The viewing distance was approximately 60 cm.

Stimuli

We used 10×10 gridded square images as “set stimuli” in our experiments as shown in Figure 1. Each of the 100 elements consisted of a small square uniformly filled with one gray scale value. There were no gaps between the adjacent squares. The 100 gray scale values within one image were randomly sampled from a Gaussian distribution with controlled mean and standard deviation values.

Two images were presented in every trial. In any single trial of both stable and unstable blocks, the two images were always identical in their mean gray scale values but different in their standard deviation values. In the stable block, the mean value of each trial was fixed to 0.5 (on a scale of 0 to 1, which is used hereafter) whereas in the unstable block the mean values of the trials were randomly set to 0.3, 0.4, 0.6, or 0.7 with an equal number of trials set at each mean value.

In every trial, one image may have *SDs* of 0.04, 0.05, and 0.06 whereas the other image had this *SD* multiplied by a gain ratio of 1.2 or 1.4. The gain ratios were determined by pilot trials to retain a moderate and flexible task difficulty. All the control variables were equally distributed within the blocks and randomly ordered.

The two images were horizontally aligned in the center of the display. Each image had a viewing angle of approximately 3.76° , and the gap between two images was 2.65° . The stimuli images were generated using Matlab R2012a (MathWorks, Natick, MA) with the gray scale values sampled from Gaussian distributions. The gray scale values were truncated to range from 0.2 to 0.8 in order to reduce extreme values. The random sampling was iterated until the error between the actual values and the preset values was less than 5%. The actual luminance of the stimuli in our display setup ranged from 3.95 to 43.85 cd/m^2 . By default, E-Prime applies a linearization function to make the gray scale values linearly related to the actual luminance. To confirm this relationship in the actual setup, we measured the gray scale–luminance transformation within the stimuli gray scale range and applied a linear regression. The results ($R^2 = 0.9844$) suggested that the transformation was close to linear in our display setup.

A mask was employed to prevent further visual processing of the stimuli images (Rolls, Tovée, & Panzeri, 1999). The same mask, a 200×200 square Gaussian noise patch with a mean value of 0.5 and *SD* of 0.23, was used throughout the experiment. The mask was displayed in the same location and in the same size as the stimuli images.

Design

We employed a 2×2 within-participant design. The independent variables were exposure time (10, 30, 50, 100, 150, 200, 300, and 1000 ms) and mean value stability (stable and unstable).

Procedure

The entire experiment included two blocks, one stable and the other unstable. The order of the blocks was balanced among the participants.

In a single trial (as illustrated in Figure 2, left panel), an initial fixation was presented for a random duration from 200 to 500 ms. The stimuli were presented for brief exposure durations, which were set to one of eight values (10, 30, 50, 100, 150, 200, 300, or 1000 ms). These eight points covered the time span within which task accuracy moves from minimum to maximum with an emphasis on the increasing phase. Pilot tests suggested that the accuracy would reach a ceiling after 1000 ms; therefore, longer durations were not tested.



Figure 1. Sample stimuli used in Experiment 1. The two images have similar mean values but different standard deviation values (in this sample, the standard deviation of the right image is greater).

The different durations were randomized within blocks. The stimuli were immediately followed by a 100-ms white noise mask.

The participants were asked to judge which of the two images exhibited higher gray scale variability and to press the corresponding key on the keyboard (“F” for left and “J” for right) as soon as they made their choice. We allowed a maximum reaction time of 2000 ms for this task. Once the subject responded, an 800-ms blank screen buffer appeared before the next trial. The participants’ responses were recorded for data analysis. Each block consisted of 192 trials (4 mean values \times 3 *SD* values \times 2 gain ratios \times 8 durations), and the entire experiment took approximately 30 min to complete.

Before the experiment, all participants underwent a practice session in which they received instructions and familiarized themselves with the experimental operations. The procedure of the practice trials was the same as in the formal experiment although the exposure time in the practice was unlimited and the correct answers were displayed simultaneously with the stimuli. In learning the relationship between the correct answers and the given stimuli, participants implicitly formed the concept of gray scale variability in their minds.

Results and discussion

The accuracy for both the stable and unstable conditions is plotted versus the exposure time in Figure 3. To better concentrate on the increasing phase, the exposure time was transformed into a base 10 logarithmic scale.

Accuracy increased with exposure time within the selected time range for both the stable and unstable conditions. The participants’ performances were consistent with chance for the extremely short exposures of 10 and 30 ms. However, with exposure times of 50 ms,

the participants were able to discriminate variability in the two images with an accuracy rate that was significantly greater than chance: stable, $t(14) = 5.53$, $p < 0.001$; unstable, $t(14) = 3.18$, $p = 0.006$. The longer exposure times enhanced participants’ performances, which achieved high accuracy levels (greater than 0.9) within 1000 ms.

A clear stable mean context advantage can be observed in Figure 3, which indicates that the accuracy of the stable blocks exceeded that of the unstable blocks. Such benefits began at 50 ms and vanished when the performance reached its maximum at 1000 ms. As Figure 3 shows, stability facilitation was more prominent in the exposure range of 100–300 ms.

The above observations were confirmed by a repeated-measures ANOVA on accuracy with the exposure time and mean context stability as two factors. The main effects of the exposure time, $F(7, 98) = 70.52$, $p < 0.001$, $\eta^2_p = 0.83$, and mean context stability, $F(1, 14) = 12.97$, $p = 0.003$, $\eta^2_p = 0.48$, were significant. The interaction of the two factors was insignificant, $F(7, 98) = 1.75$, $p = 0.107$, $\eta^2_p = 0.11$.

To better observe the stability effect, we fit the data with four-parameter sigmoid curves using the equation shown below, in which x is the log exposure time, y is the accuracy, a represents the midpoint between the top and bottom, b represents the slope of the rising phase, and “top” and “bottom” indicate the upper and lower boundaries of the accuracy, respectively. We chose sigmoid curves because they have frequently been used to model responses in cumulative processes (e.g., McKone, Martini, & Nakayama, 2001). The curves fit reasonably well, with R^2 values of 0.71 for the stable condition and 0.68 for the unstable condition.

$$y = \text{bottom} + \frac{(\text{top} - \text{bottom})}{(1 + 10^{(\log a - x) \times b})}$$

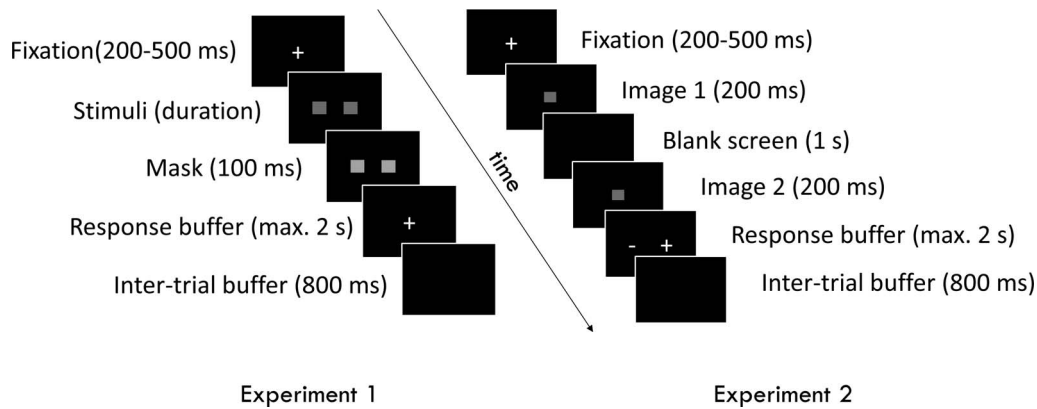


Figure 2. Schematic diagrams of the two experiments. In Experiment 1, the participants were asked to judge which of the two images had a higher gray scale variability. In Experiment 2, the participants were asked to judge whether the second image had a higher or lower gray scale variability than the first one.

We estimated the exposure time required to achieve 80% accuracy for each individual, and the t test of the estimates revealed a significant difference between the stable and unstable conditions (86 vs. 123 ms), $t(14) = 3.59$, $p = 0.003$. The stable condition had a steeper slope than the unstable condition (best fit b , 1.79 vs. 1.23), thereby suggesting that the unstable mean context slowed the growth rate of accuracy with increased processing time.

The results of Experiment 1 verified the hypothesis that the participants were able to quickly extract variability information. The immediate masking ruled out the likelihood of further processing after the stimulus display, thereby indicating that variability information was processed within a brief time, which provided evidence for the claim that ensemble variability was automatically perceived (Michael et al., 2014).

Our second hypothesis was supported by the finding that variability judgment was indeed facilitated by a stable mean context as reflected in the enhanced accuracy for stable blocks. Although the mean values were task-irrelevant, the participants performed better in stable mean contexts. This observation may be attributed to the cognitive resources saved by the repeated context (Chun & Jiang, 1998). Our results confirmed that stability facilitation could be achieved by repeated ensemble statistics (in this study, the context mean value), thus supporting the proposed functional role of ensemble statistics in building visual stability (Corbett & Melcher, 2014b). More importantly, Experiment 1 required the participants to detect the difference in a global feature (variability) rather than in local features, which were used in previous studies, suggesting that the facilitation of stable mean context is comprehensive, i.e., it covers both the local and global features in a stimuli set. Additionally, the time course of the effect suggested that the mean value stability facilitation might occur as early as 50 ms

after stimulus display and continue to affect visual processing during the 50–300 ms period.

Increased variability may result in both decreased homogeneity near the mean and increased extreme values. Observers may use either or both of these cues to encode the variability of a stimuli set. However, in certain feature domains, outliers may be automatically excluded from the computations of ensemble statistics (Haberman & Whitney, 2010). To evaluate the encoding strategies that observers may use in our task, we performed a series of analyses of the extreme values.

First, we calculated the proportion of extreme values that differs from the mean by more than 2 SD for two sets with higher and lower variability. Out of a total of 384 image pairs, the mean numbers of extreme values of the two images were small (4.45 and 4.35 out of 100 squares) and not significantly different from one another, $t(383) = 0.95$, $p = 0.34$.

We also calculated the kurtosis and skewness of the gray scale values of each image, which partially reflected the severity of extreme values. A higher kurtosis indicates that more of the variability is due to

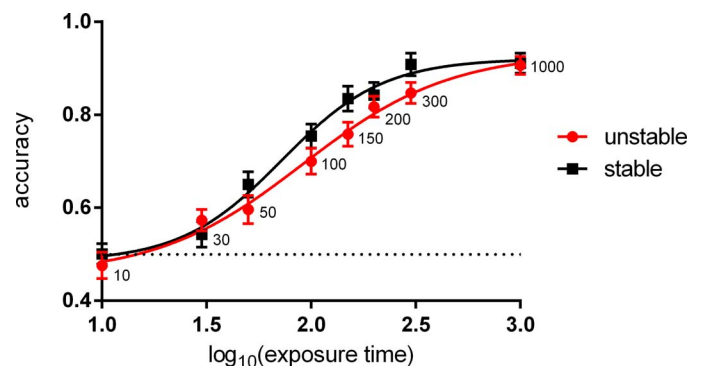


Figure 3. Accuracy as a function of exposure time in Experiment 1 ($n = 15$). Error bars represent the standard error of the mean values. The dotted line represents the chance level.

	Block 1	Block 2	Block 3	Block 4
Between-trials	stable	unstable	stable	unstable
Within-trial	stable	stable	unstable	unstable
Mean value	.5	.3/.4/.6/.7	.5	.3/.4/.6/.7
Mean ratio	1	1	.8/1/1.2	.8/1/1.2
SD value	.04/.05/.06	.04/.05/.06	.04/.05/.06	.04/.05/.06
SD ratio	.7/.9/1.1/1.3	.7/.9/1.1/1.3	.7/.9/1.1/1.3	.7/.9/1.1/1.3

Table 1. Mean and standard deviation values in the four blocks in Experiment 2.

several extreme differences from the mean values, and the kurtosis can increase while the standard deviation remains the same if more of the variation is caused by extreme values. There was no significant difference in the kurtosis between the paired images with higher and lower variabilities in each trial (2.97 vs. 2.93), $t(383) = 1.39$, $p = 0.17$.

A positive skewness means that the extreme values greater than the mean are farther from the mean than extreme values less than the mean and vice versa. There was no significant skewness difference between the paired images with higher versus lower variability in each trial, $t(383) = 0.78$, $p = 0.44$. The skewness was small (0.019 vs. 0.006) and not different from 0, $t(383) = 1.53$, $p = 0.13$, 0.59 .

To evaluate the effect of extreme values on accuracy, we performed regression analyses with the number of extreme values as independent factors. The results revealed that the number of extreme values did contribute to the accuracy, $F(1, 382) = 4.78$, $p = 0.03$; however, the contribution was only 1% (R^2). Our regressions on differences in skewness and kurtosis revealed that these characteristics did not contribute to accuracy, $F(1, 382) = 0.11$, 0.54 , $p = 0.75$, 0.46 .

To summarize, the analyses of extreme values suggested that the weight of outliers in variability perception was low in our task. We inferred that observers encode in a more holistic manner, thus emphasizing the general homogeneity of the stimuli near mean values.

Experiment 2

Experiment 1 demonstrated that a stable mean value context facilitated the task of variability judgment; however, which characteristics of the variability judgment are affected remains unclear. The manipulation of mean stability in Experiment 1 was straightforward but failed to capture the multiple levels on which the stability effect may occur. For instance, repeated features of successive stimuli within trials may provide stability in a transient sense whereas repeated features between trials may induce a longer-term stability effect across the entire experiment block. Because there was only one

“general” stability manipulated in Experiment 1, it was impossible to dissect the stability effect in finer detail.

Experiment 2 thus aimed to investigate the above issues using a successive display paradigm in which the context mean stability was manipulated by two independent variables, i.e., within-trial stability and between-trials stability. In each trial, the participants were presented with two successive images and asked to judge whether the second image had higher or lower variability than the first one. A within-subject 2×2 block design was adopted to investigate the effects of those “substabilities.” We analyzed the participants’ sensitivity and response bias to the variability information under the influence of different mean contexts.

Methods

Participants

A total of 15 naïve participants (eight females; mean age = 23 years; $SD = 2$ years) were recruited from nearby universities. All the participants had normal or corrected-to-normal vision. All the participants received monetary rewards for completing the experiment. The participants in Experiment 2 were not involved in Experiment 1.

Apparatus

Same as Experiment 1.

Stimuli

The general features of the stimuli images were the same as those in Experiment 1. The stimuli image pairs were categorized into four types (as shown in Table 1) according to the 2 (within-trial stable/unstable) \times 2 (between-trials stable/unstable) experimental design.

In the within-trial stable blocks, the two images of the same trial always had equal mean values whereas in the within-trial unstable blocks the mean values of the second images were the mean values of the first image multiplied by a gain ratio, which might be 0.8, 1, or 1.2. In the between-trials stable blocks, the mean values of the first images in each trial were fixed at 0.5 whereas in

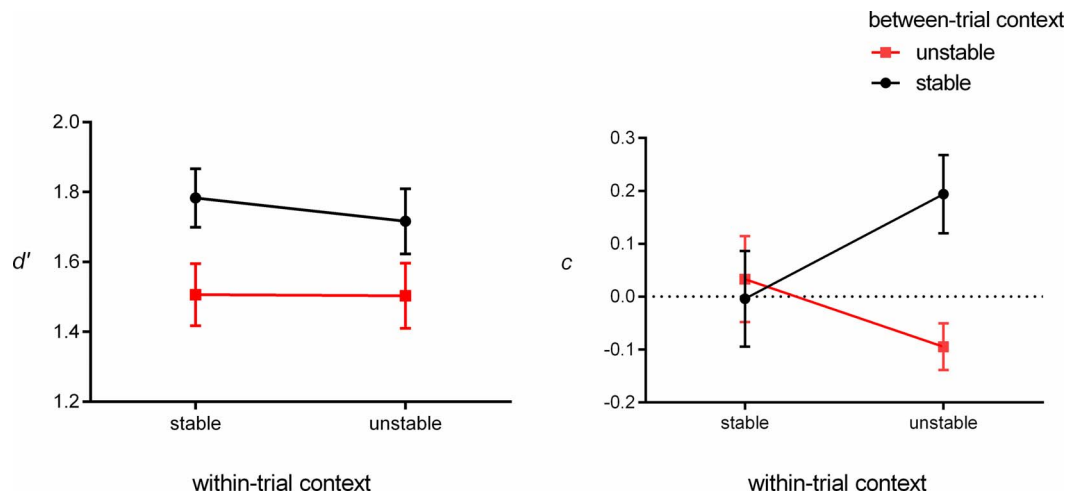


Figure 4. d' and c values of the four stability conditions in Experiment 2 ($n = 15$). The error bars represent the standard error of the mean. The dotted line in the right panel indicates unbiased c .

the between-trials unstable blocks the mean values of the first images were varied and took a value of 0.3, 0.4, 0.6, or 0.7.

The standard deviations were controlled as in Experiment 1. The standard deviation of the first image could be 0.04, 0.05, or 0.06, and the standard deviation of the second image was the value of the first multiplied by a gain ratio, which could be 0.7, 0.9, 1.1, or 1.3. In this manner, the second images could have lower or higher standard deviation values. All the control variables were equally distributed in a random order.

Design

We employed a 2×2 within-participant design. The independent variables were the between-trials mean stability and the within-trial mean stability.

Procedure

The entire experiment included four blocks (within-trial stable/unstable by between-trials stable/unstable). The order of the blocks was balanced among the participants.

In a single trial (as illustrated in Figure 2, right panel), after an initial fixation presented for a random duration of 200 to 500 ms, two successive images were presented for 200 ms each with an interstimulus interval (ISI) of 1000 ms of blank screen. In determining the exposure time, our goal was to allow the participants sufficient time to process the images. The results of Experiment 1 demonstrated that the correct rate of variability judgment within 200 ms was greater than 80%. Although a longer exposure time might improve performance further, the improvement was limited, and it would have taken much more time to complete the experiment. The 1000-ms ISI was chosen

to clearly separate the two stimuli, which made it less likely to integrate the two ensembles.

The participants were asked to judge whether the second image increased or decreased in variability compared with the first image. They were asked to press the corresponding key (“F” for decrease and “J” for increase) on the keyboard as soon as they made their choice. A maximum reaction time of 2000 ms was allowed. After the response, an 800-ms blank screen buffer would appear before the next trial. No mask was added because both the exposure time and interim time in Experiment 2 were long, which reduced the benefit of precise time control. The participants’ responses and accuracy were recorded for additional analyses. Each block consisted of 144 trials (4 mean values \times 3 mean ratios \times 3 SD values \times 4 SD ratios). The entire experiment took approximately 45 min to complete. The settings for the practice sessions were similar to those in Experiment 1.

Results and discussion

We calculated d' and c from the response data (Macmillan & Creelman, 2005). A correct “decrease” response was denoted as a “hit,” and an incorrect “decrease” response was denoted as a “false alarm.” In the present study, d' measured the participants’ sensitivity to variability information, and c measured the shift of the response criterion. A larger d' indicates higher sensitivity, and a positive c indicates a bias toward underestimation.

As shown in the left panel of Figure 4, stable mean values between trials enabled the participants to perform with higher sensitivity in stable between-trials conditions as indicated by a larger d' . The effect of within-trial mean stability on the variability judgment

task was less significant. Repeated measures ANOVA on d' confirmed the above observations with a significant main effect of between-trials mean stability (1.75 vs. 1.51), $F(1, 14) = 13.34$, $p = 0.003$, $\eta^2_p = 0.49$, and an insignificant main effect of within-trial mean stability (1.64 vs. 1.61), $F(1, 14) = 0.29$, $p = 0.595$, $\eta^2_p = 0.02$. The interaction of the two factors was insignificant, $F(1, 14) = 0.373$, $p = 0.551$, $\eta^2_p = 0.03$.

Criterion c exhibited a different pattern from that of d' . The right panel of Figure 4 shows that the two blocks with stable within-trial contexts had c values near zero, $ts(14) = -0.044, 0.41$, $ps = 0.965, 0.688$, thus indicating no bias for either response whereas the two blocks with unstable within-trial mean context were biased toward the opposite directions as determined by the stability of the between-trials mean context. For unstable within-trial mean context, one sample t test from zero indicated that a stable between-trial context led the participants to underestimate the variability, $t(14) = 2.62$, $p = 0.02$, whereas an unstable between-trials context led the participants to overestimate the variability, $t(14) = -2.15$, $p = 0.049$. Repeated measures ANOVA on c confirmed the above observations with a significant main effect of between-trials mean stability, $F(1, 14) = 8.93$, $p = 0.01$, $\eta^2_p = 0.49$, and an insignificant effect of within-trial mean stability, $F(1, 14) = 0.58$, $p = 0.46$, $\eta^2_p = 0.04$. The interaction of the two factors was also significant, $F(1, 14) = 13.51$, $p = 0.002$, $\eta^2_p = 0.49$. A simple effect analysis showed that the biases of stable and unstable between-trials contexts were comparable for the stable within-trial condition, $F(1, 14) = 0.26$, $p = 0.617$, $\eta^2_p = 0.02$, but differed from one another for the unstable within-trial condition, $F(1, 14) = 39.03$, $p < 0.001$, $\eta^2_p = 0.74$.

The sensitivity was lower in the context of varying mean values between trials. This effect may reflect the additional cognitive load from the continuously changing mean contexts. Although the mean values were task-irrelevant in our experiment, their stability nevertheless captured attentional resources and influenced the primary task. However, the within-trial stability of mean values had little influence on participants' sensitivity in the variability task. This finding is notable because in most of the trials from the within-trial unstable blocks the participants were required to encode the first image and compare its variability with that of the second image, which had a different mean value. This varying transient mean context was expected to result in greater cognitive load and less sensitivity, but this impairment did not occur. One possible explanation may be that the difference between the two images from the same trial was highly relevant to the task, and participants focused more attentional resources in this transient context as a result. Therefore, they might have been aware of the changing within-trial mean context. Because they understood that mean values are task-irrelevant, they

might have intentionally resisted the distortion from the unstable mean context.

Previous studies have suggested that humans tend to underestimate the variability of the visual environment (Kareev, Arnon, & Horwitz-Zeliger, 2002). Our results suggested that underestimation occurred in a “globally stable” although “locally unstable” scenario: We may underestimate variability when context mean values change in a transient manner but when the mean context remains constant over the long term. However, we may also overestimate variability when context mean values change in a “globally unstable” and “locally unstable” scenario. Bauer (2009) has suggested that observers may overestimate mean size when mean size varies between trials. These results indicated that unstable mean context can bias mean perception as well as variability perception and that the direction of the bias is determined by the interaction between long-term stability and transient mean stability.

As in Experiment 1, we conducted control analyses of extreme values. Of the total of 576 image pairs, the mean proportions of extreme gray scale values of two images were small (4.47 vs. 4.38 out of 100 squares) and not significantly different from one another, $t(575) = 0.95$, $p = 0.34$. There was no significant kurtosis difference between the paired images with larger and smaller variability, $t(575) = 0.93$, $p = 0.35$. There was no skewness difference between the images with larger and smaller variability, $t(575) = 0.03$, $p = 0.98$. The skewness was small (-0.0003 vs. -0.0008) and not significantly different from 0, $t(575) = -0.04, -0.07$, $p = 0.97, 0.94$. Regression analyses with a number of extreme values, skewness difference, and kurtosis difference as independent factors revealed that extreme values did not contribute to the accuracy, $F(1, 574) = 0.02, 1.13, 0.02$, $p = 0.90, 0.29, 0.88$.

General discussion

The present study demonstrated that the stability of contextual mean values influences the perception of variability by affecting the observer's sensitivity and response criterion. The loss of sensitivity is primarily attributed to an unstable between-trials mean context whereas the response bias is determined by the interaction between the within-trial stability and the between-trials mean stability.

Stable mean context facilitates variability perception

There are two primary paradigmatic differences between the present study and Corbett and Melcher

(2014b), who first introduced the effect of statistical stability facilitation. The first is the manipulation of stability. In Corbett and Melcher (2014b), stability was built and then changed. In our study, we built and maintained stability in the stable blocks, but stability was never built in the unstable blocks. The second difference is the task type. Whereas visual search is by nature a task based on local features of a stimulus set, the variability judgment used in our study is by nature a task based on the global features of a set.

Despite these differences, there is a fundamental similarity between the two studies. The “stability facilitation effect” in both studies is conceptually the same and can be described as “performance enhancement after being visually presented with stimuli with identical global features in context,” which is why we could discuss our results on the basis of the stability effect identified by Corbett and Melcher (2014b).

How do ensemble statistics maintain visual stability to facilitate ongoing perception? Corbett and Melcher (2014b) suggested that statistical stability might function by freeing attentional resources. This claim implied a potential hypothesis that monitoring unstable statistical representation requires additional attention. The results of Experiment 1 support this hypothesis by revealing an advantage of stable over unstable mean context in processing variability. The early onset of stability facilitation and the rapid perception of variability in Experiment 1 also suggest that ensemble statistics may be extracted automatically. Taken together, these observations indicate that although global properties can be automatically extracted, monitoring changes in such properties requires additional attention. Recent work has suggested that distributed attention may be responsible for such a requirement (Bajjal, Nakatani, van Leeuwen, & Srinivasan, 2013).

In Experiment 2, we found that the statistical stability of the context could be categorized on different levels and that these “substabilities” may function differently in perceptual decisions. This finding is consistent with previous research that found that ensemble statistics were computed and stored on multiple levels (Corbett & Melcher, 2014a). In our study, the stability was confined to global features (context mean values), and we divided this stability into between-trials and within-trial levels. This division allowed us to view the functional role of ensemble statistics not only through a “local versus global” perspective but also in a “transient versus long-term” manner.

Our results suggested that between-trials—but not within-trial—mean stability modulates sensitivity to variability. We speculate that the influence on sensitivity may be an implicit process. By contrast, the participants may have been more prone to perceiving

the within-trial transient context change, which might lead to intentional neglect. Participants’ concentration on variability information prevented sensitivity losses; however, the variability information presented in different mean values nevertheless affected their criterion of choice. More generally, our results imply that transient changes in global context features may result in biased estimates of the attended features whereas lower-frequency changes can implicitly impair our perceptual sensitivity.

Peterson and Beach (1967) used the term “intuitive statistician” to suggest that humans have a natural sensitivity to statistical parameters. This idea was revised by Juslin, Winman, and Hansson (2007), who used the term “naïve intuitive statistician,” positing that even when our minds are sensitive to the statistical properties of the given data, we are frequently biased when making decisions by inference. Our results suggest that when the statistical context is unstable, individuals are more prone to making biased judgments, and the bias is due to the weakened sensitivity and shifted criterion in the noisy environment. Therefore, the stability of the statistical context affects both sampling (by modulating sensitivity) and inference (by biasing criterion) in decision making.

Encoding variability

Observers may use one or more statistical parameters to encode variability. Previous studies have suggested that candidates may be standard deviation (Bex & Makous, 2002; Moulden, Kingdom, & Gatley, 1990) or range (Ariely, 2001; Lovie, 1978) of the set. The outlier analysis in the present study suggest that the extreme values in the set might not be a critical feature that observers used—at least in our tasks—to access variability information. This result may undermine the candidacy of range as the parameter for variability perception because of its high susceptibility to outliers. The fact that observers use the similarity of group members near the mean to encode variability suggests that variability perception may rely on mean perception.

Notably, another possibility is that humans do not have fixed statistical parameters for variability perception. Beach and Scopp (1968) found that participants might accurately compare the relative variability between two groups of stimuli but could not accurately estimate the absolute variability of one stimulus set. Observers may neither understand nor use statistical formulas to compute variability but might instead develop the ability to compare variability as a result of adaptation (Pollard, 1984). In this manner, the “formula” used by the observers may depend on their individual experience.

The relationship between mean and variability perception

Previous studies have suggested that extraction of ensemble statistics may be an automatic process because mean representation has a rapid processing speed (Chong & Treisman, 2003). Experiment 1 supported this claim from the perspective of variability perception. The results of Experiment 1 suggested that the extraction of variability information is rapid: Participants' accuracy in discriminating the variability differences reached a high level (greater than 0.75) in less than 100 ms, which indicates the rapid encoding of mean information (Chong & Treisman, 2003).

Is the similarity of the processing speed of mean and variability information a consequence of the two parts sharing a common processing mechanism? The Garner (1974) interference paradigm for facial expression/identity may provide useful insight on the question of separate or shared mechanism. Regarding the analysis in Garner, if the processing of a task-relevant feature is not influenced by the values of another task-irrelevant feature, then we can infer that the processing of these two features is separate. Similarly, if mean and variability perception are separate, we would expect that the variability judgment would be unaffected by the stability of mean context, but our results suggest that such is not the case. The impairment of the variability task due to unstable mean contexts indicates that there is competition between mean and variability processing, which would be a consequence of the hypothesized shared mechanism.

Our results echo the findings of the recent work regarding the “priming by variance” phenomenon, in which the matching of the prime and target variability facilitates extraction of the mean representation even when the variability is task-irrelevant (Michael et al., 2014). The “priming by variance” phenomenon demonstrated the impact of variability processing on mean representation whereas our study demonstrated the impact of mean representation on variability processing. Taken together, these results suggest that the processing of mean and variability information may not be totally separate.

However, we have not yet determined the specific stage(s) at which the two processes compete and interact with one another. Does the competition occur at the stage when attention is captured or during encoding? Our results suggest both albeit with different weights. With regard to the manner in which attention is captured, competition may affect explicit and implicit attention differently, such that unstable transient mean context (within-trial) may arouse explicit attention to resist reduced sensitivity to variability whereas the unstable longer-term context (between-trials) may

impair variability sensitivity by implicitly capturing attention.

The influence on the encoding stage is more complicated. Our outlier analysis implies that participants mainly rely on the dispersion of values near the mean, suggesting that the extraction of mean value may play a role in variability processing. However, imaging studies provide preliminary evidence that the neural substrates for mean and variability may be different: the anterior-medial ventral visual cortex might play an important role in forming averages in humans (Cant & Xu, 2012) whereas the lateral geniculate nucleus may be responsible for variability processing in cats (Bonin, Mante, & Carandini, 2006). We would not recommend combining human and cat studies to draw any conclusions; nevertheless, these two studies suggest that the encoding of mean and variability information may occur in different areas of the brain and at different stages of visual processing.

Such attempts to combine human and animal studies are indicative of the scarcity of imaging studies on ensemble statistics. We mentioned that Cant and Xu (2012) used the fMRI adaption paradigm to locate brain areas that are responsible for mean representations in humans and found that the anterior–medial ventral visual cortex might play an important role in forming averages of stimulus sets. However, there is no equivalent fMRI study of variability representations. Brenner, Bialek, and de Ruyter van Steveninck (2000) demonstrated that changes in the variability information of a stimulus ensemble might arouse adaptive rescaling of the input/output function in the visual system of the blowfly. Once again, it is impossible to make a direct analog to the human visual system although this study provides hints that we might look into assessing the variability information as an indicator of the dynamic range of visual input. Michael et al. (2014) also suggested that the variability of visual information helps to set the gain of neural processing during perceptual choice. From this perspective, the finding that variability modulated the performance for the mean value extraction task (Marchant et al., 2013) is easily understood.

Notably, no study to date has prevented such extraction of summary statistics. Few, if any, studies have used patients with brain lesions as participants or used techniques that could alternate the brain function of healthy participants (e.g., transcranial magnetic stimulation or transcranial direct current stimulation) to locate the neural substrate of central tendency and variability coding. Future studies may take advantage of such methods to reveal the brain mechanism that underlies ensemble statistics and the interplay between mean and variability perception.

Keywords: ensemble statistics, contextual stability, variability perception

Acknowledgments

This study was supported by grants from the National Basic Research Program of China (2011CB302201) and the National Natural Science Foundation of China (31371031). We thank the anonymous reviewers and the editor for their insightful and constructive comments, which were very helpful for us and led to substantial improvements in this manuscript.

Commercial relationships: none.

Corresponding author: Wenfeng Chen.

Email: chenwf@psych.ac.cn.

Address: State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China.

References

- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122–131.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12(2), 157–162.
- Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & Psychophysics*, 70(7), 1325–1326.
- Baijal, S., Nakatani, C., van Leeuwen, C., & Srinivasan, N. (2013). Processing statistics: An examination of focused and distributed attention using event related potentials. *Vision Research*, 85, 20–25.
- Bauer, B. (2009). The danger of trial-by-trial knowledge of results in perceptual averaging studies. *Attention, Perception, & Psychophysics*, 71(3), 655–665.
- Beach, L. R., & Scopp, T. S. (1968). Intuitive statistical inferences about variances. *Organizational Behavior and Human Performance*, 3(2), 109–123.
- Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America A*, 19(6), 1096–1106.
- Bonin, V., Mante, V., & Carandini, M. (2006). The statistical computation underlying contrast gain control. *The Journal of Neuroscience*, 26(23), 6346–6353.
- Brenner, N., Bialek, W., & de Ruyter van Steveninck, R. (2000). Adaptive rescaling maximizes information transmission. *Neuron*, 26(3), 695–702.
- Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *Journal of Neuroscience*, 32(22), 7685–7700.
- Chong, S. C., Joo, S. J., Emmanouil, T.-A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics*, 70(7), 1327–1334.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43(4), 393–404.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36(1), 28–71.
- Corbett, J. E., & Melcher, D. (2014a). Characterizing ensemble statistics: Mean size is represented across multiple frames of reference. *Attention, Perception, & Psychophysics*, 76(3), 746–758.
- Corbett, J. E., & Melcher, D. (2014b). Stable statistical representations facilitate visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 40(5), 1915–1925.
- Dakin, S. C., Mareschal, I., & Bex, P. J. (2005). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vision Research*, 45(24), 3027–3049.
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics*, 72(7), 1825–1838.
- Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review*, 114(3), 678–703.
- Kareev, Y., Arnon, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General*, 131(2), 287–297.
- Lovie, P. (1978). Teaching intuitive statistics II. Aiding the estimation of standard deviations. *International Journal of Mathematical Education in Science and Technology*, 9(2), 213–219.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica*, 142(2), 245–250.

- McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology. Human Perception and Performance*, 27(3), 573–599.
- Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences, USA*, 111(21), 7873–7878.
- Morgan, M. J., Mareschal, I., Chubb, C., & Solomon, J. A. (2012). Perceived pattern regularity computed as a summary statistic: Implications for camouflage. *Proceedings of the Royal Society B: Biological Sciences*, 279(1739), 2754–2760.
- Moulden, B., Kingdom, F., & Gatley, L. F. (1990). The standard deviation of luminance as a metric for contrast in random-dot images. *Perception*, 19(1), 79–101.
- Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, 70(5), 772–788.
- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68(1), 29–46.
- Pollard, P. (1984). Intuitive judgments of proportions, means, and variances: A review. *Current Psychology*, 3(1), 5–18.
- Rolls, E. T., Tovée, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience*, 11(3), 300–311.
- Schweickert, R., Han, H. J., Yamaguchi, M., & Fortin, C. (2014). Estimating averages from distributions of tone durations. *Attention, Perception, & Psychophysics*, 76(2), 605–620.
- Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 329–337.
- Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention ensemble perception of a crowd's gaze. *Psychological Science*, 25(10), 1903–1913.
- Tong, K., Tang, W., Chen, W., & Fu, X. (2015). Statistical summary representation: Contents and mechanisms. *Advances in Psychological Science (in Chinese)*, 23(10), 1723–1731.
- Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, 15(2), 77–84.