Quality of average representation can be enhanced by refined individual items

Jihong Lee¹ · Sang Chul Chong^{1,2}

Accepted: 6 September 2020 / Published online: 8 October 2020
© The Psychonomic Society, Inc. 2020

Abstract

Ensemble perception is efficient because it summarizes redundant and complex information. However, it loses the fine details of individual items during the averaging process. Such characteristics of ensemble perception are similar to those of coarse processing. Here, we tested whether extracting an average of a set was similar to coarse processing. To manipulate coarse processing, we used the fast flicker adaptation known as suppressing coarse information processed by the magnocellular pathway. We hypothesized that if computing the average of a set relied on coarse processing, the precision of an averaging task should decrease after adaptation compared to baseline (no-adaptation). Across experiments with various features (orientation in Experiment 1, size in Experiment 2, and facial expression in Experiment 3), we found that suppressing coarse information did not disrupt the performance of the averaging tasks. Rather, adaptation increased the precision of mean representation. The precision of mean representation might have increased because fine information was relatively enhanced after adaptation. Our results suggest that the quality of ensemble representation relies on that of individual items.

Keywords Ensemble perception · Averaging · Coarse processing · Fast flicker adaptation

Introduction

Economical management of resources is a virtue not only in our ordinary life but also in cognitive life. Humans efficiently manage their limited resources using various strategies. For example, imagine that you were a farmer waiting for grapes to ripen and trying to determine when to begin harvesting. You might check the sizes and colors of grapes immediately in front of you. However, if you were facing a vast vineyard, looking wider and seeing the overall sizes and colors of the vineyard would be a better strategy than focusing on a specific bunch of grapes. This cognitive management is known as *ensemble perception*, a strategy of managing cognitive resources economically. To summarize multiple stimuli in a scene, it is helpful to distribute one's attention over them rather than focusing on them individually (Baek & Chong, 2020a,

2020b; Chong & Evans, 2011; Chong & Treisman, 2005; Treisman, 2006). Using this strategy, people can rapidly extract summary statistics of multiple stimuli such as mean and variance (Chong & Treisman, 2003; Morgan, Chubb, & Solomon, 2008). This mode is efficient because it helps us to grasp the gist of a scene rapidly (Alvarez, 2011; Cohen, Dennett, & Kanwisher, 2016; Hochstein, 2019; Hochstein, Pavlovskaya, Bonne, & Soroker, 2015).

Traditionally, perceiving the gist of a scene has been understood in the framework of the coarse-to-fine strategy (Hegdé, 2008). Similar to the grapes example, people usually process visual information in a scene coarsely at first, and in fine detail afterward. In other words, we rapidly obtain a gist by perceiving a scene coarsely, and, if necessary, scrutinize details later. This coarse-to-fine mechanism is operated via two separate pathways in charge of spatiotemporally different visual information (Merigan & Maunsell, 1993). The magnocellular (M-) pathway is sensitive to temporal information and is less sensitive to spatial information, whereas the parvocellular (P-) pathway is sensitive to spatial information and less sensitive to temporal information. Hence, coarse processing is more closely related to the M-pathway, through which blurry and rapid visual inputs including low spatial frequencies (LSF) are mainly projected, whereas fine processing is related to the P-pathway, through which sharp and slow

✉ Sang Chul Chong
scchong@yonsei.ac.kr

¹ Graduate Program in Cognitive Science, Yonsei University, Seoul, Republic of Korea

² Department of Psychology, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

inputs including high spatial frequencies (HSF) are projected. Previous studies have suggested that LSF information rapidly forms a scene-gist through the M-pathway and, through the feedback connection, facilitates the follow-up finer analysis based on HSF information through the P-pathway (Bar, 2003; Bar et al., 2006; Bognár et al., 2017; Goffaux et al., 2011; Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Petras, ten Oever, Jacobs, & Goffaux, 2019; Schyns & Oliva, 1994).

In the framework of the coarse-to fine strategy, ensemble perception has characteristics that are similar to those of coarse processing for gist perception. First, ensemble perception is fast like coarse processing. Previous studies showed that people can perceive ensemble information of a set in a very brief time, i.e., as little as 50 ms (mean size: Chong & Treisman, 2003; mean emotion: Haberman & Whitney, 2009; Li et al., 2016; mean lifelikeness: Yamanashi Leib, Kosovicheva, & Whitney, 2016). Moreover, this ensemble information is quite robust even when the quality of individual items is low. For example, when multiple items are too crowded to be discriminated, people can still extract ensemble information from them (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Likewise, patients with unilateral spatial neglect can extract average information from the neglected field, although they cannot extract individual information from it (Pavlovskaya, Soroker, Bonneh, & Hochstein, 2015; Yamanashi Leib, Landau, Back, Chong, & Robertson, 2012a; for a review, see Hochstein et al., 2015). Similar results were obtained in patients with prosopagnosia (Yamanashi Leib, Puri, et al., 2012b).

Fast and robust processing of ensemble information results in a summary of information with reduced detail. While we can rapidly obtain summary information from an ensemble, we lose specific details of individual items in the process of averaging. For example, Ariely (2001) found that after a set of spots with various sizes was displayed, participants could not differentiate members of the presented set from non-members, but could discriminate between their mean sizes. Alvarez and Oliva (2008), with a multiple-object-tracking task, showed that participants had little information about the location of an unattended item, but they could report the centroid of multiple unattended items. Finally, Haberman and Whitney (2011) demonstrated that when two sets of 16 faces were successively displayed, participants showed above-chance performance of discriminating between mean facial expressions of two sets even in trials where they failed to localize changes in individual items. This representation of overall properties without access to details leads to an assumption that ensemble processing is similar to coarse processing. One possible explanation for these characteristics is that when people represent multiple items, they extract summary statistics not by using fine information of individual items but by using only coarse information.

Nevertheless, an ensemble representation may be formed based on fine information as well as coarse information. Previous studies suggest that the precision of ensemble representation relies on the quality of the individual items comprising a set. For example, Jacoby et al. (2013) showed that masked individual items contributed less to forming an ensemble representation. Likewise, Sun and Chong (2020) demonstrated that an increasing number of inverted faces in a set decreased the accuracy of a mean facial expression discrimination task. Given that the quality of visual inputs is related to fine information, the quality of summary may be obtained from fine processing through the P-pathway. Furthermore, Brady, Shafer-Skelton, and Alvarez showed that spatial layout information can be grasped even without LSF information (Brady, Shafer-Skelton, & Alvarez, 2017). They found that participants' performance of rapid-scene recognition was related to their sensitivity to changes in spatial ensemble structures comprised of HSF Gabors. Overall, these results suggest that ensemble information is not merely coarse processing dissociated from fine information of individuals.

In the current study, we investigated whether ensemble perception relies on coarse information (i.e., only LSF information). To this end, we tested whether averaging performance could be disrupted by suppressing the M-pathway. We adopted a fast flicker adaptation paradigm used by Arnold et al. (2016). They used rapidly flickering dynamic noise patches to alter the M-pathway. Using this method, it is possible to selectively suppress M-pathway inputs without affecting P-pathway inputs. In their study, they presented participants with hybrid face images (mixtures of a blurred feminine face and a sharpened masculine face) and asked them to judge the gender of images. The result showed that participants' responses were biased towards the sharpened gender images after adaptation, suggesting that the flicker adaptation made them rely on fine information from the P-pathway. In addition, participants showed better performances on the tests relying on HSF information (i.e., Vernier acuity and fine print acuity tests) after adaptation. Specifically, they showed that the contrast thresholds of detecting an LSF Gabor increased after adaptation, whereas those of detecting an HSF Gabor were unchanged. These results suggest that the presentation of rapidly flickering dynamic noise patches selectively suppresses the M-pathway.

Here, we adapted the M-pathway using the fast flicker adaptation paradigm and tested whether it influenced the precision of averaging. If ensemble perception relied on only coarse information, adapting to fast flickering noise would reduce the precision of averaging. To test this hypothesis, we investigated whether adaptation affected the precision of averaging orientations in Experiment 1, sizes in Experiment 2, and facial expressions in Experiment 3.

Experiment 1: Mean orientation judgment task

In Experiment 1, we tested whether the flicker adaptation influenced averaging orientations. To this end, we measured and compared participants' precision of discriminating mean orientations depending on adaptation. We also added a heterogeneity condition, where a set with the same number of stimuli had a variance (heterogeneous) or not (homogeneous). If averaging relies on coarse processing, averaging performance should be disturbed by reduced M-pathway information.

Methods

Participants

Fifteen subjects ($M_{age} = 26.8$ years, $SD_{age} = 3.02$; eight males) including the first author participated in Experiment 1. We determined the sample size using comparable data from Arnold et al. (2016), where the effects of adapting the M-pathway were shown. In their second experiment, 16 participants were used to compare estimated thresholds of the Vernier acuity test, and a large-sized adaptation effect ($\eta_p^2 = .31$) was observed. We conducted a simulation-based power analysis for a two-way within-subject repeated-measures ANOVA on the basis of statistical reports in the previous study. From 10,000 iterations of the simulation using generated data, we estimated statistical power of .80 for the main effect of adaptation based on a Type 1 error of .05. All participants reported normal or corrected-to-normal visual acuity. The experimental procedures were approved by the Yonsei University Institutional Review Board. Written informed consent was obtained from all participants prior to the commencement of the study. After the completion of experiments, they were paid 5,000 KRW per 30 min, except the first author.

Apparatus and stimuli

All stimuli were generated using the Psychophysics Toolbox Version 3 extension (Brainard, 1997; Pelli, 1997) for MATLAB (MathWorks Inc., Natick, MA, USA) and were displayed on a gamma-corrected CRT monitor (HP P1230) with a refresh rate of 85 Hz. All experiments were conducted in a dark room, and participants were seated at 60 cm away from the monitor with their head on a head-chin rest.

For mean orientation judgments, 60 Gabor patches were used for a probe display. Each Gabor was a sine wave grating of five cycles per degree (cpd), enveloped by a Gaussian filter with a standard deviation of 7 min of arc. The mean test angles had seven levels (tilted either -4.5° , -3.0° , -1.5° , 0.0° , 1.5° , 3.0° , or 4.5° from the vertical meridian). In every trial, 60 Gabors were presented in a circular area with diameters of

3.5° of visual angle (dva), where each position was randomly sampled without overlaps. In the homogeneous condition, each of the 60 Gabors had the same orientation with a test mean orientation, whereas in the heterogeneous condition each Gabor orientation was sampled from a uniform distribution and normalized to a test mean orientation and a standard deviation of 4° . For adaptation, we adopted the method of Arnold et al. (2016), in which a participant was exposed for a few seconds to a dynamic noise pattern updated at the monitor refresh rate. For each trial, we made as many noise patterns as there were frames in the period of adaptation (85 patterns per second) and presented them sequentially. In each frame, the noise pattern was a grid whose individual cells had sides of 0.2 dva and a luminance level randomly sampled from a normal distribution with a mean of 127 (background luminance) and a standard deviation of 50 on a 0–255 luminance scale. To ensure that the adaptation area had the same shape as the testing area for orientation judgments, noise patterns were trimmed in a circular shape. The adaptation area covered by the adapter was 1.32 times larger than that of probe displays, to maximize the effect of adaptation. Throughout the experiment, a black crosshair was presented around stimuli as a reference frame, which turned red in each response phase.

Design and procedure

The experiment was conducted as a blocked design to test 2 (adaptation) $\times$ 2 (heterogeneity) conditions within a participant. For each condition, seven levels of mean orientation were repeated 24 times and presented in a pseudo-randomly intermixed way. Each session consisted of four blocks corresponding to each condition (A: adaptation-homogeneous, B: no-adaptation-homogeneous, C: adaptation-heterogeneous, and D: no-adaptation-heterogeneous). The block sequence was randomized following a Latin square design, where adaptation blocks were not in succession (i.e., A-B-C-D, B-A-D-C, C-D-A-B, and D-C-B-A). Participants were assigned to one of the four block sequences, and they were presented four blocks per session, and they performed an additional session on a different day with a reverse sequence compared to the first day (e.g., B-A-D-C $\rightarrow$ C-D-A-B). The total number of trials was 672.

The experimental procedure is depicted in Fig. 1. In the adaptation block, the trial started with adaptation in which a flickering dynamic noise adaptor was presented for 10 s in the first trial, and for 5 s for subsequent trials. After adaptation, there was a 250-ms inter-stimulus interval (ISI) and a test display was presented for 250 ms. In the no-adaptation block, the trial started without adaptation, but there was a blank display for 1 s before the test display appeared. After the test display disappeared, the reference frame turned red and participants were asked to indicate whether Gabors were tilted on average clockwise (CW) or counterclockwise (CCW) with

Fig. 1 Procedure of Experiment 1

keyboard arrow keys (right for CW and left for CCW). At the end of each block, a 3-min break was provided to prevent adaptation effects from getting carried over into the next. Before the main session, participants carried out practice trials to be familiarized with the average orientation discrimination task. In the practice session, participants completed 14 trials for each heterogeneity condition without adaptation, and auditory feedback was provided when responses were incorrect. There was no feedback in the main session to prevent the formation of bias (Bauer, 2009).

Analysis

Using *Palamedes* version 1.10.3 (Prins & Kingdom, 2018), the proportion of CW responses to the seven test orientations was fitted to a cumulative Gaussian function for each condition. To measure sensitivity to the difference between the reference (0°) and test stimuli ($\pm 1.5^\circ$, $\pm 3.0^\circ$, and $\pm 4.5^\circ$), we used the slope parameter estimated from a fitted psychometric function, where a larger parameter value indicated higher sensitivity. We obtained four slopes corresponding to each condition from each individual. A two-way repeated-measures ANOVA was conducted on the slopes for each condition (adaptation $\times$ heterogeneity). Additionally, we conducted a two-way Bayesian repeated-measures ANOVA on the same data. By using inclusion Bayes factors ($BF_{\text{inclusion}}$) based on matched models, we observed how much the inclusion of each main effect and interaction was supported by the data. For example, $BF_{\text{inclusion}} = 5$ means that a model with such a term is five times more likely to explain the data than a model without that term, whereas a $BF_{\text{inclusion}} = 0.10$ ($BF_{\text{exclusion}} = 10$) means that exclusion of that term is ten times likely to explain the data.

Both frequentist and Bayesian analyses were performed using JASP Version 0.10.2 (JASP Team, 2019).

Results and discussion

The average slopes of each condition are plotted in Fig. 2. We investigated whether the flicker adaptation influenced averaging and whether this effect varied depending on the heterogeneity of a set. We found that participants were more sensitive to mean orientation after adaptation than at baseline (no-adaptation), and they were more sensitive when orientations were

Fig. 2 Results of Experiment 1. The proportion of “clockwise” responses to the test angles of each condition was fitted to a cumulative Gaussian function at the individual level. The slope of the fitted function was a parameter indicating the sensitivity, where a large slope indicates high sensitivity. Dark gray bars indicate average slopes of the no-adaptation condition (baseline), whereas light gray indicates those in the adaptation condition. Error bars denote the 95% within-subject confidence interval

homogeneous than when orientations were heterogeneous. Furthermore, we found that increase in sensitivity after adaptation in the homogeneous condition was larger than that in the heterogeneous condition. These conclusions were supported by the following statistical analyses. We observed a significant main effect of adaptation, $F(1, 14) = 59.70$, $p < .001$, $\eta_p^2 = .81$, $BF_{\text{inclusion}} = 21.52$, showing a larger slope for discriminating mean orientation after adaptation ($M = 0.95$, $SD = 0.31$) than baseline (no-adaptation, $M = 0.51$, $SD = 0.11$). Participants also showed a larger slope in the homogenous condition ($M = 0.81$, $SD = 0.39$) than in the heterogeneous condition ($M = 0.65$, $SD = 0.21$), $F(1, 14) = 20.21$, $p < .001$, $\eta_p^2 = .59$, $BF_{\text{inclusion}} = 6.64 \times 10^8$. We also observed a significant interaction between the adaptation and heterogeneity conditions, $F(1, 14) = 13.38$, $p = .003$, $\eta_p^2 = .49$, $BF_{\text{inclusion}} = 30.65$. Post hoc comparisons showed that contrasts between the adaptation and no-adaptation conditions were significant under both, homogenous ($M_{\text{diff}} = 0.60$, $SE = 0.68$, $t(14) = 8.93$, $p_{\text{bonf}} < 0.001$) and heterogeneous conditions ($M_{\text{diff}} = 0.27$, $SE = 0.72$, $t(14) = 3.76$, $p_{\text{bonf}} = 0.005$). Furthermore, contrasts between the homogeneous and heterogeneous conditions were significant under the adaptation condition ($M_{\text{diff}} = 0.33$, $SE = 0.06$, $t(14) = 5.68$, $p_{\text{bonf}} < 0.001$), but not under the no-adaptation condition ($M_{\text{diff}} = 0.00$, $SE = 0.06$, $t(14) = 0.04$, $p_{\text{bonf}} = 1.000$). These results suggest that the increase in the slope after adaptation was larger in the homogeneous condition than in the heterogeneous condition.

We found that the precision of averaging orientations did not deteriorate in either the homogenous or the heterogeneous conditions after exposure to the flicker adaptor, suggesting that averaging orientations was not solely reliant on M-pathway information. Rather, adaptation increased the precision of averaging in both the heterogeneity conditions. This effect was similar to the improved performance in the tests relying on HSF information in Arnold et al. (2016), suggesting that more precise representations of visual inputs were helpful to discriminate mean orientation. Although the adaptation effect in terms of precision improvements was larger in the homogeneous condition than in the heterogeneous condition, the effect was still significant in the heterogeneous condition. Thus, these results suggest that adaptation does not interfere with the averaging process and that the precision of mean orientation relies on fine information of visual inputs.

Experiment 2: Mean size judgment task

In Experiment 1, we found that the flicker adaptation did not deteriorate but increased the precision of averaging orientations. To generalize the results of Experiment 1, we tested whether the adaptation influenced the precision of averaging sizes. To this end, we measured and compared participants' precision of discriminating mean sizes of two sets, depending

on the adaptation. Additionally, we presented stimuli in the periphery to increase the influence of LSF information compared to Experiment 1 where the stimuli were presented around the fovea.

Methods

Participants

Twelve subjects ($M_{\text{age}} = 25.5$ years, $SD_{\text{age}} = 1.88$; six males) participated in Experiment 2. Among them, 11, including the first author, took part in Experiment 1. We determined this sample size based on the results of Experiment 1. Similar to Experiment 1, we conducted a simulation-based power analysis for a two-way repeated-measures ANOVA. The estimated power was 1.00 for the main effect of adaptation, .95 for the heterogeneity effect, and .85 for their interaction based on a Type 1 error of .05. All participants reported normal or corrected-to-normal visual acuity. The experimental procedures were approved by the Yonsei University Institutional Review Board and informed written consent was obtained from all participants before participation. After participation, they were paid 5,000 KRW per 30 min, except for the author.

Stimuli

Fourteen circles were used for a mean size discrimination task. Each outlined circle had a line width of 0.05 dva. Each circle was located at the virtual vertices of two concentric polygons comprised of an inner hexagon and outer octagon, whose vertices deviated 2 dva and 4 dva from the center, respectively. To avoid using the same positions, in every trial each polygonal array rotated to a random extent, and each position of vertices was randomly jittered up to 0.3 dva horizontally and vertically. The test mean diameter had seven levels (0.85, 0.90, 0.95, 1.00, 1.05, 1.10, and 1.15 dva). In the homogeneous condition, all circles had the same diameter with a test mean diameter, whereas in the heterogeneous condition, each circle diameter was sampled from a uniform distribution and normalized to a test mean diameter and a standard deviation of 0.18 dva. Reference array always had the mean diameter of 1.00 dva, and test array had one of seven test mean diameters. In each test display, reference and test arrays were located at the left or right fields, respectively, and their center had an eccentricity of 8 dva from the central fixation cross, and their fields were randomly interchanged in every trial. For adaptation, general methods were the same as in Experiment 1, except dual adaptors were used to cover each area where the reference and test arrays were presented. Each adaptation area was 1.56 times larger than the area of stimuli array to maximize the effect of adaptation.

Design and procedure

The experimental design was the same as in Experiment 1. General procedures in Experiment 2 were the same as in Experiment 1 except stimuli in the adaptation and test display were presented on both sides of the central fixation cross (depicted in Fig. 3). Participants were instructed to fixate on the fixation cross during both the adaptation and test phases. After the test stimuli disappeared and the fixation cross turned red, they reported which visual field had larger circles on average with the left and right arrow keys on a keyboard. Before each main session, a practice session with auditory feedback was provided as in Experiment 1. There was no feedback during the main session.

Analysis

As in Experiment 1, we obtained slopes of each condition by fitting the proportion of “larger” responses on test stimuli to a cumulative Gaussian function. We conducted two-way repeated-measures ANOVAs by both frequentist and Bayesian approaches on slopes with 2 (adaptation) $\times$ 2 (heterogeneity) conditions.

Results and discussion

The average slopes of each condition are plotted in Fig. 4. We found that the sensitivity of averaging was higher in the homogeneous than in heterogeneous conditions. We found marginally increased changes in sensitivity after adaptation, and these changes did not vary depending on the heterogeneity condition. These were supported by the following statistical

Fig. 4 Results of Experiment 2. The proportion of “larger” responses to the test set of each condition was fitted to a cumulative Gaussian function at the individual level. Dark gray bars indicate average slopes of the no-adaptation condition (baseline) and light gray bars indicate that of the adaptation condition. Error bars denote the 95% within-subject confidence interval

analyses. We observed a significant main effect of heterogeneity, $F(1, 11) = 19.69$, $p < .001$, $\eta_p^2 = .64$, $BF_{\text{inclusion}} = 665.77$, showing the larger slope in the homogeneous ($M = 19.01$, $SD = 5.87$) than in the heterogeneous conditions ($M = 12.90$, $SD = 5.29$). The main effect of adaptation was not significant in the frequentist analysis, $F(1, 11) = 4.64$, $p = .054$, $\eta_p^2 = .30$. However, results based on a Bayesian approach showed that the inclusion of the adaptation factor was more likely than exclusion ($BF_{\text{inclusion}} = 2.22$), indicating a marginally larger slope in the adaptation ($M = 17.42$, $SD = 7.29$) compared to the no-adaptation conditions ($M = 14.49$, $SD = 4.92$). The interaction between the adaptation and

Fig. 3 Procedure of Experiment 2

heterogeneity conditions was not significant, $F(1, 11) = 0.00$, $p = .959$, $\eta_p^2 = .00$, $BF_{\text{inclusion}} = 0.38$.

Unlike Experiment 1, we presented stimuli in the periphery to increase the reliance of the task on LSF information. Nevertheless, we did not find any reduction in the precision of discriminating mean sizes after flicker adaptation in both heterogeneity conditions as in Experiment 1. We found weak evidence for incremental changes in the precision of averaging after adaptation. We do not think that this is because of the location of stimuli (8 dva away from the central fixation). Arnold et al. (2016) showed that the adaptation effect of enhancing spatial acuity was not different depending on eccentricity, although increasing eccentricity reduced the spatial acuity. However, peripheral presentation of stimuli could have made items more crowded because we used multiple items unlike Arnold et al. (2016). The increased spatial uncertainty due to crowding might have offset the influence of sharpened individual items after adaptation. Nevertheless, we did not observe any reduction in precision after adaptation, despite the increased influence of coarse information by increasing eccentricity. Thus, we can at least conclude that averaging could be performed when coarse information was suppressed, consistent with the results of Experiment 1.

Experiment 3: Mean facial expression judgment task

In Experiments 1 and 2, we found that the flicker adaptation did not decrease the precision of averaging orientations or sizes. In Experiment 3, we tested whether the adaptation influenced averaging more complex features, such as facial expressions. In this experiment, we measured and compared participants' precisions of discriminating mean facial expressions depending on the adaptation. As in Experiments 1 and 2, we included a heterogeneity condition. In addition, we used two types of emotion, happy and angry, to test whether the adaptation effect differed depending on emotion.

Methods

Participants

Fourteen subjects ($M_{\text{age}} = 25.57$ years, $SD_{\text{age}} = 1.80$; seven males) participated in Experiment 3. This sample size was calculated on the basis of the results of Experiments 1 and 2. Similar to Experiments 1 and 2, we conducted a simulation-based power analysis for a three-way repeated-measures ANOVA, and the estimated power was 1.00 for the main effect of adaptation, .99 for the heterogeneity effect, and .61 for their interaction based on a Type 1 error of .05. Two participants also participated in Experiment 1, and 10 participants, including the first author, participated in both

Experiments 1 and 2. Because one participant dropped out after the second session, 13 participants completed the entire experiment. All participants reported normal or corrected-to-normal visual acuity. The experimental procedures were approved by the Yonsei University Institutional Review Board, and we obtained informed written consent from all participants prior to participation. After the experiments were completed, they were paid 5,000 KRW per 30 min, except the first author.

Stimuli

For a mean facial expression judgment task, we utilized a morphed face stimuli set used by Sun and Chong (2020). This stimulus set comprised two vectors relevant to emotion types (happy and angry) and each vector included 201 faces. A vector represents an emotional scale based on the norm-based coding account of facial expressions (Gwinn, Matera, O'Neil, & Webster, 2018; Palermo et al., 2018). Each scale ranges from the full-blown expression to the anti-emotion, and the intensity of the emotion was expressed as the extent to which each facial expression deviates from a face with a neutral expression. Specifically, the intensity of each emotion was defined on the basis of the ratio of full-blown to neutral expression in a morphed face. Each vector includes a neutral face (0%) and 100 emotional faces ranging from 1% to 100% (full-blown). In addition, this vector includes 100 levels (-100% to -1%) of anti-emotion faces by morphing faces from the neutral face to the opposite direction of the emotional faces. Thus, the same intensity levels of emotional and anti-emotional faces (e.g., 50% and -50%) correspond to the same physical differences from the neutral face, but they were symmetrical to each other with respect to the neutral face. In a mean facial expression judgment task, eight oval-shaped faces were used. Each face subtended to 1.46×1.92 dva² and was located at the virtual vertices of two concentric squares. Each vertex deviated 1.5 dva for the inner square and 3 dva for the outer one from the center of the screen. In every trial, an inner square-shaped array rotated to a random degree, and an outer array rotated 40–50° more than the inner one. Each position of vertices was randomly jittered up to 0.1 dva horizontally and vertically to prevent the stimuli from being presented in a too regularly. The test facial expression intensity had seven levels (0%, 13%, 26%, 39%, 52%, 65%, and 78%). In the homogeneous condition, all faces had the same value with a test mean intensity. In the heterogeneous condition, each intensity of individual faces was sampled from a uniform distribution and normalized to a test mean intensity and a target standard deviation of 10% as much as possible. Note that the intensity of facial expressions should be expressed as an integer because it was limited to discrete index numbers of the face vector (one level change in an index corresponded to 1% change in the intensity). Thus, we first randomly sampled

eight numbers from a uniform distribution, normalized them to a test mean and a target standard deviation (10%) and rescaled them as follows: first, we rounded up eight sampled numbers and averaged them. If the mean of the rounded numbers was not the same as the test mean, then we randomly picked one of the rounded numbers and adjusted it such that the mean of the rounded numbers matched the test mean as far as possible. The reference array always had a mean intensity of 0%, and the test array had one of seven test mean intensities. Please note that the anti-emotional faces (-1% to -100%) were not included in the test means because we found that sensitivity to anti-emotional faces was lower than emotional faces in the results of a pilot study, consistent with results in Sun and Chong (2020). Nevertheless, individual faces in a heterogeneous block (particularly in lower test mean displays, such as 0% and 13%) were sometimes sampled from an anti-emotional vector. For example, for a given trial with a test mean of 0%, individual faces might be -17%, -10%, -6%, 1%, 4%, 7%, 9%, and 12% ($M = 0\%$, $SD = 10.11$). For adaptation, general methods were the same as in Experiments 1 and 2. Each adaptation area was 2.89 times larger than the area of stimuli array to maximize the effect of adaptation.

Design and procedure

The experimental design was the same as in Experiments 1 and 2 except that participants participated in four sessions because the two emotions were used (happy and angry).

General procedures in Experiment 3 were the same as those in Experiments 1 and 2, except that the reference and test stimuli were sequentially presented (depicted in Fig. 5). In the adaptation phase, a flickering dynamic noise adaptor was

presented for 10 s in the first trial, and for 5 s in subsequent trials. After adaptation, the reference and test stimuli were serially presented for 250 ms each with an ISI of 500 ms, and their order was randomly decided in every trial. When face stimuli disappeared, and the crosshair turned red, participants were asked to indicate which interval (former or latter) had faces with a stronger facial expression on average with keyboard number pads (1 to the former and 2 to the latter). Before each main session, a practice session with auditory feedback was provided as in Experiments 1 and 2. One difference was that a practice session was repeated up to three times because participants reported the task of averaging facial expressions difficult in a pilot experiment. Feedback was not provided in the main session.

Analysis

As in Experiments 1 and 2, the proportion of “stronger” responses was fitted to a cumulative Gaussian function. Note that we used the slope in Experiments 1 and 2 – because it is a parameter inversely proportional to the differential threshold – to identify test levels that were larger or smaller from the reference level. By comparison, test levels in Experiment 3 were always the same or larger than the reference level and the chance level was 50%. Thus, we obtained the threshold where the function reached 75% and defined a reciprocal of a threshold as the sensitivity of each condition. Thus, the sensitivity in Experiment 3 was inversely proportional to the absolute threshold in order to detect the presence of average expression against neutral expressions (the reference) at the chance level. Therefore, sensitivity in Experiment 3 can also be considered

Fig. 5 Procedure for Experiment 3

a parameter denoting the precision of mean representation, like the slopes in Experiments 1 and 2.

We conducted both frequentist and Bayesian versions of three-way repeated-measures ANOVAs on sensitivities with a 2 (adaptation) $\times$ 2 (heterogeneity) $\times$ 2 (emotion) design.

Results and discussion

The average sensitivities of each condition are plotted in Fig. 6. We examined whether the adaptation influenced averaging facial expressions, and the pattern of effect changed depending on the heterogeneity and the emotion type of a set. We found that participants were more sensitive to mean facial expression after adaptation than without adaptation. In addition, they were more sensitive to sets of happy faces than those of angry faces. Sensitivity was not different between the homogeneous and heterogeneous conditions, and the pattern of adaptation was not different depending on either heterogeneity or emotion. These results were supported by following statistical analyses. We observed a significant main effect of adaptation, $F(1, 12) = 6.82$, $p = .023$, $\eta_p^2 = .36$, $BF_{\text{inclusion}} = 40.78$, showing the higher sensitivity for discriminating mean facial expressions after adaptation ($M = 6.46 \times 10^{-2}$, $SD = 2.22 \times 10^{-2}$) than at baseline (no-adaptation, $M = 5.29 \times 10^{-2}$, $SD = 1.83 \times 10^{-2}$). Participants showed higher sensitivity to happy faces ($M = 6.95 \times 10^{-2}$, $SD = 2.11 \times 10^{-2}$) than to angry faces ($M = 4.80 \times 10^{-2}$, $SD = 1.47 \times 10^{-2}$), $F(1, 12) = 23.19$, $p < .001$, $\eta_p^2 = .66$, $BF_{\text{inclusion}} = 2.83 \times 10^6$. However, sensitivity was not significantly different between the homogeneous ($M = 6.04 \times 10^{-2}$, $SD = 1.79 \times 10^{-2}$) and heterogeneous conditions ($M = 5.71 \times 10^{-2}$, $SD = 2.39 \times 10^{-2}$), $F(1, 12) = 1.13$, $p = .309$,

$\eta_p^2 = .01$, $BF_{\text{inclusion}} = 0.29$. In addition, we did not observe any significant interaction among conditions, emotion type $\times$ adaptation: $F(1, 12) = 0.54$, $p = 0.475$, $\eta_p^2 = .04$, $BF_{\text{inclusion}} = 0.34$; emotion type $\times$ heterogeneity: $F(1, 12) = 2.90$, $p = .114$, $\eta_p^2 = .19$, $BF_{\text{inclusion}} = 0.97$; adaptation $\times$ heterogeneity: $F(1, 12) = 0.55$, $p = .473$, $\eta_p^2 = .04$, $BF_{\text{inclusion}} = 0.31$; emotion type $\times$ adaptation $\times$ heterogeneity: $F(1, 12) = 0.36$, $p = .561$, $\eta_p^2 = .03$, $BF_{\text{inclusion}} = 0.44$, indicating that the pattern of adaptation effect did not differ depending on the other conditions.

Consistent with those of averaging orientations and sizes in Experiments 1 and 2, the performance of averaging facial expressions did not decrease after the flicker adaptation. Rather, we found that the precision of averaging increased after adaptation, and this pattern did not vary depending on the heterogeneity or emotion of a set. These results suggest that reducing LSF information helps to form a more precise mean representation. Furthermore, the adaptation effect did not differ depending on different emotions, suggesting that the ability to average facial expressions might be similar across emotions. Similarly, Sun and Chong (2020) showed that the influence of individual faces on mean facial expression did not differ depending on different emotions, although perceiving individual emotions relied on different facial features (e.g., the eyes for angry or the mouth for happy faces). In addition, they showed that the increasing number of inverted faces gradually disrupted the precision of averaging facial expressions, suggesting that the precision of individual faces is related to that of averaging them. Overall, our results suggest that ensemble computation does not rely solely on coarse information based on the M-pathway but might also be based on the fine information of visual inputs.

General discussion

Representing summary statistics of multiple items is an efficient strategy, but at the same time, it results in the loss of the fine details of the included items. These characteristics of ensemble perception therefore have properties similar to those of coarse processing. We thus tested whether ensemble processing was similar to coarse processing based on LSF information through the M-pathway. To this end, we used the fast flicker adaptation method (Arnold et al., 2016) to suppress the M-pathway where coarse information is mainly processed. Consistently across our three experiments, we found that reducing the contribution of LSF information by adaptation did not decrease the precision of averaging multiple items. Instead, we observed that reducing coarse information from visual inputs improved the quality of their summary representation. Arnold et al. (2016) suggested that the flicker adaptation increased spatial acuity, because fine processing was relatively facilitated by reduced coarse information after adaptation. In the current study, we similarly found that the flicker

Fig. 6 Results of Experiment 3. The proportion of “stronger” responses to the test facial expression of each condition was fitted to a cumulative Gaussian function at the individual level. We defined the sensitivity as a reciprocal of the threshold estimated from the fitted function. Dark gray bars indicate average sensitivities of the no-adaptation condition (baseline) and light gray bars indicate those of the adaptation condition. Error bars denote the 95% within-subject confidence interval

adaptation improved the precision of mean representation. In Experiment 1, we observed that adaptation helped to form a more precise mean orientation. In Experiment 3, adaptation also increased the precision of averaging facial expressions. Even in Experiment 2, we found marginal increments in the precision of averaging circle sizes after adaptation. These results suggest that the flicker adaptation improves the quality of mean representation.

We found a similar adaptation effect across all three experiments although the stimuli were presented in different ways – at the center of vision in Experiment 1, side by side in the periphery in Experiment 2, and successively at the center in Experiment 3. We do not think the adaptation effects are a result of the difference in stimulus locations or ways of presentation, because the stimuli were presented identically in each experiment depending on the adaptation conditions. Similarly, Arnold et al. (2016) showed that adaptation effects do not differ depending on stimulus eccentricity. Despite the differences in presentation methods across experiments, we think that adaptation effects on the quality of mean representation are generalized to various feature types because of the similar patterns of results that we found regardless of the type of averaging.

We did not find that suppressing coarse processing disturbed ensemble processing, despite their similarities. This might be because the averaging process involves two stages (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013; Baek & Chong, 2020a; Parkes et al., 2001; Solomon, 2010; Solomon, Morgan, & Chubb, 2011): encoding individual items and averaging them. The flicker adaptation might have only influenced the stage of encoding individual items without affecting averaging. We think that blurry LSF information might function like noise in the encoding stage because it provides blurry and sketchy information. This early noise degrades individual representations, and in turn, reduces the precision of ensemble representation. The flicker adaptation might have reduced early noise involved with encoding individual items by reducing blurriness and coarseness of visual inputs. The reduced noise in individual representations may have then made the mean representations more precise.

This noise reduction hypothesis by adaptation is consistent with the previous findings (Jacoby et al., 2013; Sun & Chong, 2020). Jacoby et al. (2013) found that individual items with less visibility due to object substitution masking contributed less to mean representation. Likewise, Sun and Chong (2020) found that the increased number of inverted faces in a set degraded the accuracy of discriminating mean facial expressions. Because individual face processing was disturbed by the inversion, adding an inverted face to a set increased noise for individual expressions to be averaged. The added noise reduced the accuracy of the mean computation. Together with these findings, our results suggest that the quality of mean representation relies on that of individual ones.

Our results suggest that ensemble representation is robust to the effects of early noise in the encoding stage because of noise cancellation during the averaging process, consistent with the results of previous studies (Baek & Chong, 2020a; Lee, Baek, & Chong, 2016; Sun & Chong, 2020). We found that the flicker adaptation increased the precision of mean representation in both heterogeneity conditions, similarly in Experiments 2 and 3. However, in Experiment 1, the increased precision of mean representation by adaptation was smaller in the heterogeneous condition than in the homogenous condition. We think that this difference might be due to noise cancellation occurring at the averaging stage (Alvarez, 2011; Baek & Chong, 2020a; Jacoby et al., 2013; Sun & Chong, 2020). Early noise in individual items can be gradually canceled out with an increasing number of items to be averaged (Baek & Chong, 2020a; Lee et al., 2016; Parkes et al., 2001; Robitaille & Harris, 2011). If a set size is large enough, early noise can be less influential on the quality of ensemble representation because early noise is canceled out during the averaging process. Thus, the adaptation benefit of reducing early noise can be overshadowed by noise cancellation in averaging with a large set size. Since the set size was much larger in Experiment 1 (60) than the other two experiments (14 in Experiment 2 and eight in Experiment 3), the adaptation benefit might not have been as influential in Experiment 1 due to large noise cancellation. This might be the reason why the adaptation benefit was smaller in the heterogeneous condition of Experiment 1. People might use fewer items in the averaging task with a homogeneous set than with a heterogeneous set, because the task could also be performed using a single item, rather than multiple items. This would reduce noise cancellation when averaging a homogeneous set compared to a heterogeneous set. Therefore, our results suggest that the quality of ensemble representation relies not only on the quality of individual items but also on the number of individual items.

In the current study, we found that the precision of mean representation was improved, rather than impaired, by suppressing the M-pathway, suggesting that ensemble representation can be precisely formed based only on HSF information. Our results suggest that suppressing the P-pathway should reduce the quality of ensemble representation. To our knowledge, however, there is no known method to adapt the P-pathway selectively. Although we did not test the effect of P-pathway suppression on ensemble perception in this study, our results suffice to show the contribution of the P-pathway to ensemble processing because we replicated Arnold and colleagues' findings of improved performances on the tests relying on HSF information after adaptation (Arnold et al., 2016). Nevertheless, our results do not suggest that ensemble processing relies only on the fine information of individual items. Rather, we think that LSF information can also contribute to ensemble processing. In terms of precision, the visual system seems to be ordinarily in a *suboptimal* state where

intact LSF information probably deteriorates the quality of visual representation (Arnold et al., 2016). However, coarse processing is beneficial because LSF information through the M-pathway is rapidly processed compared to fine information through the P-pathway. We think that ensemble perception adaptively utilizes LSF and HSF information depending on available resources. Previous studies showed that people could compute mean representation even with very short exposure time, but to some extent, could compute it more precisely when the exposure time increased (Chong & Treisman, 2003; Haberman & Whitney, 2009; Li et al., 2016). In particular, Li et al. (2016) showed that as exposure time increased, response time to discriminate mean facial expressions increased, and, furthermore, the accuracy also increased to some degree. These results suggest that ensemble representation can be formed with coarse sketches under time pressure. However, when enough time to process individuals is given, more precise ensemble representation is possible. Thus, even less precise and coarse processing could be helpful to form ensemble representations rapidly, especially when processing time is limited. In our study, we only tested a single presentation duration of 250 ms, which was enough time to form ensemble representations based on fine information through the P-pathway. Future studies are needed to investigate whether suppressing LSF information would disturb the formation of ensemble representation when processing time is limited.

Conclusion

Our study demonstrates that suppressing coarse information through the M-pathway does not decrease but increases the precision of mean representation. These results support the notion that ensemble perception is not confined to coarse processing based on LSF information but is based on detailed information of individual items. Thus, the quality of ensemble representation depends on that of individual items.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2019R1A2B5B01070038). Data in this research article were previously presented at the 19th annual meeting of the Vision Sciences Society (May 2019).

Open practices statement The raw data for all experiments are available on the Open Science Framework (<https://osf.io/dwrtq>), and none of the experiments was preregistered.

References

- Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research*, 83, 25–39. <https://doi.org/10.1016/j.visres.2013.02.018>
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122–131. <https://doi.org/10.1016/j.tics.2011.01.003>
- Alvarez, G. A., & Oliva, A. (2008). The Representation of Simple Ensemble Visual Features Outside the Focus of Attention. *Psychological Science*, 19(4), 392–398. <https://doi.org/10.1111/j.1467-9280.2008.02098.x>
- Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science*, 12(2), 157–162. <https://doi.org/10.1111/1467-9280.00327>
- Arnold, D. H., Williams, J. D., Phipps, N. E., & Goodale, M. A. (2016). Sharpening vision by adapting to flicker. *Proceedings of the National Academy of Sciences*, 113(44), 12556–12561. <https://doi.org/10.1073/pnas.1609330113>
- Back, J., & Chong, S. C. (2020a). Distributed attention model of perceptual averaging. *Attention, Perception, and Psychophysics*, 82(1), 63–79. <https://doi.org/10.3758/s13414-019-01827-z>
- Back, J., & Chong, S. C. (2020b). Ensemble perception and focused attention: Two different modes of visual processing to cope with limited capacity. *Psychonomic Bulletin & Review, Advance on*, 1–5. <https://doi.org/10.3758/s13423-020-01718-7>
- Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, 15(4), 600–609. <https://doi.org/10.1162/089892903321662976>
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Schmidt, A. M., ... Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0507062103>
- Bauer, B. (2009). The danger of trial-by-trial knowledge of results in perceptual averaging studies. *Attention, Perception, and Psychophysics*. <https://doi.org/10.3758/APP.71.3.655>
- Bognár, A., Csete, G., Németh, M., Csibri, P., Kincses, T. Z., & Sári, G. (2017). Transcranial stimulation of the orbitofrontal cortex affects decisions about magnocellular optimized stimuli. *Frontiers in Neuroscience*. <https://doi.org/10.3389/fnins.2017.00234>
- Brady, T. F., Shafer-Skelton, A., & Alvarez, G. A. (2017). Global ensemble texture representations are critical to rapid scene perception. *Journal of Experimental Psychology: Human Perception and Performance*, 43(6), 1160–1176. <https://doi.org/10.1037/xhp0000399>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*. <https://doi.org/10.1163/156856897X00357>
- Chong, S. C., & Evans, K. K. (2011). Distributed versus focused attention (count vs estimate). *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6), 634–638. <https://doi.org/10.1002/wcs.136>
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43(4), 393–404. [https://doi.org/10.1016/S0042-6989\(02\)00596-5](https://doi.org/10.1016/S0042-6989(02)00596-5)
- Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception and Psychophysics*, 67(1), 1–13. <https://doi.org/10.3758/BF03195009>
- Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the Bandwidth of Perceptual Experience? *Trends in Cognitive Sciences*, 20(5), 324–335. <https://doi.org/10.1016/j.tics.2016.03.006>
- Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From Coarse to Fine? Spatial and Temporal Dynamics of Cortical Face Processing. *Cerebral Cortex*, 21(2), 467–476. <https://doi.org/10.1093/cercor/bhq112>
- Gwinn, O. S., Matera, C. N., O'Neil, S. F., & Webster, M. A. (2018). Asymmetric neural responses for facial expressions and anti-expressions. *Neuropsychologia*, 119, 405–416. <https://doi.org/10.1016/j.neuropsychologia.2018.09.001>

- Haberman, J., & Whitney, D. (2009). Seeing the Mean: Ensemble Coding for Sets of Faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718–734. <https://doi.org/10.1037/a0013899>
- Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review*, 18(5), 855–859. <https://doi.org/10.3758/s13423-011-0125-6>
- Hegd, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, 84(4), 405–439. <https://doi.org/10.1016/j.pneurobio.2007.09.001>
- Hochstein, S. (2019). The gist of Anne Treisman’s revolution. *Attention, Perception, and Psychophysics*, 1–7. <https://doi.org/10.3758/s13414-019-01797-2>
- Hochstein, S., Pavlovskaya, M., Bonne, Y. S., & Soroker, N. (2015). Global statistics are not neglected. *Journal of Vision*, 15(4), 7. <https://doi.org/10.1167/15.4.7>
- Jacoby, O., Kamke, M. R., & Mattingley, J. B. (2013). Is the whole really more than the sum of its parts? Estimates of average size and orientation are susceptible to object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 233–244. <https://doi.org/10.1037/a0028762>
- JASP Team. (2019). JASP (Version 0.10.2). [Computer Software].
- Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, 107, 49–57. <https://doi.org/10.1016/j.vision.2014.11.013>
- Lee, H., Baek, J., & Chong, S. C. (2016). Perceived magnitude of visual displays: Area, numerosity, and mean size. *Journal of Vision*, 16(3), 12. <https://doi.org/10.1167/16.3.12>
- Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X. (2016). Processing of Individual Items during Ensemble Coding of Facial Expressions. *Frontiers in Psychology*, 7, 1332. <https://doi.org/10.3389/fpsyg.2016.01332>
- Merigan, W. H., & Maunsell, J. H. R. (1993). How Parallel are the Primate Visual Pathways. *Annual Review of Neuroscience*, 16(1), 369–402. <https://doi.org/10.1146/annurev.neuro.16.1.369>
- Morgan, M., Chubb, C., & Solomon, J. A. (2008). A “dipper” function for texture discrimination based on orientation variance. *Journal of Vision*, 8(11), 9–9. <https://doi.org/10.1167/8.11.9>
- Palermo, R., Jeffery, L., Lewandowsky, J., Fiorentini, C., Irons, J. L., Dawel, A., ... Rhodes, G. (2018). Adaptive face coding contributes to individual differences in facial expression recognition independently of affective factors. *Journal of Experimental Psychology: Human Perception and Performance*, 44(4), 503–517. <https://doi.org/10.1037/xhp0000463>
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739–744. <https://doi.org/10.1038/89532>
- Pavlovskaya, M., Soroker, N., Bonne, Y. S., & Hochstein, S. (2015). Computing an average when part of the population is not perceived. *Journal of Cognitive Neuroscience*, 27(7), 1397–1411. https://doi.org/10.1162/jocn_a_00791
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <https://doi.org/10.1163/156856897X00366>
- Petrs, K., ten Oever, S., ten Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, 186(April 2018), 103–112. <https://doi.org/10.1016/j.neuroimage.2018.10.086>
- Prins, N., & Kingdom, F. A. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Frontiers in Psychology*, 9(JUL), 1–14. <https://doi.org/10.3389/fpsyg.2018.01250>
- Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision*, 11(12), 1–8. <https://doi.org/10.1167/11.12.18>
- Schyns, P. G., & Oliva, A. (1994). From Blobs to Boundary Edges : Evidence for Time- and Spatial-Scale-Dependent Scene Recognition, 5(4), 195–200.
- Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, 10(14), 19–19. <https://doi.org/10.1167/10.14.19>
- Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, 11(12), 13–13. <https://doi.org/10.1167/11.12.13>
- Sun, J., & Chong, S. C. (2020). Power of Averaging: Noise Reduction by Ensemble Coding of Multiple Faces. *Journal of Experimental Psychology: General*, 149(3), 550–563. <https://doi.org/10.1037/xge0000667>
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, 14(4–8), 411–443. <https://doi.org/10.1080/13506280500195250>
- Yamanashi Leib, A., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, 7, 1–10. <https://doi.org/10.1038/ncomms13186>
- Yamanashi Leib, A., Landau, A. N., Baek, Y., Chong, S. C., & Robertson, L. (2012a). Extracting the mean size across the visual field in patients with mild, chronic unilateral neglect. *Frontiers in Human Neuroscience*, 6(October), 1–11. <https://doi.org/10.3389/fnhum.2012.00267>
- Yamanashi Leib, A., Puri, A. M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012b). Crowd perception in prosopagnosia. *Neuropsychologia*, 50(7), 1698–1707. <https://doi.org/10.1016/j.neuropsychologia.2012.03.026>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.