Comparing averaging limits for social cues over space and time

Joseph Florey Experimental Psychology, Queen Mary University of London, London, UK

Steven C. Dakin School of Optometry and Vision Science, University of Auckland, Auckland, New Zealand
UCL Institute of Ophthalmology, University College London, London, UK

Isabelle Mareschal Experimental Psychology, Queen Mary University of London, London, UK

Observers are able to extract summary statistics from groups of faces, such as their mean emotion or identity. This can be done for faces presented simultaneously and also from sequences of faces presented at a fixed location. Equivalent noise analysis, which estimates an observer's *internal noise* (the uncertainty in judging a single element) and *effective sample size* (ESS; the effective number of elements being used to judge the average), reveals what limits an observer's averaging performance. It has recently been shown that observers have lower ESSs and higher internal noise for judging the mean gaze direction of a group of spatially distributed faces compared to the mean head direction of the same faces. In this study, we use the equivalent noise technique to compare limits on these two cues to social attention under two presentation conditions: spatially distributed and sequentially presented. We find that the differences in ESS are replicated in spatial arrays but disappear when both cue types are averaged over time, suggesting that limited peripheral gaze perception prevents accurate averaging performance. Correlation analysis across participants revealed generic limits for internal noise that may act across stimulus and presentation types, but no clear shared limits for ESS. This result supports the idea of some shared neural mechanisms b in early stages of visual processing.

Introduction

Observers' ability to extract summary statistics from groups of objects is well established. The perceived mean of low-level properties such as orientation (Dakin, 2001; Dakin & Watt, 1997; Solomon, 2010), size (Ariely, 2001; Chong & Treisman, 2005b) and motion (Dakin, Mareschal, & Bex, 2005) can be reliably estimated from groups (*ensemble stimuli*). Recently it has been demonstrated that these summary statistics can be estimated over complex, "higher level" properties such as facial emotion, identity, and gaze direction (Florey, Clifford, Dakin, & Mareschal, 2016; Haberman & Whitney, 2009; Sweeny & Whitney, 2014). Although most research has focused on averaging across spatially distributed arrays of items, observers can also average over temporal sequences (Albrecht, Scholl, & Chun, 2012; Gorea, Belkoura, & Solomon, 2014; Haberman, Harp, & Whitney, 2009; Piazza, Sweeny, Wessel, Silver, & Whitney, 2013). We have previously shown that observers' averaging of gaze direction over space is limited compared to averaging of head direction (Florey et al., 2016). This sets an important limit on our ability to process crowds of faces, because it has been shown that humans are more sensitive to the direction of attention of a group of faces than they are to an individual face (Gallup et al., 2012). The question remains, however, whether there is a difference in how well people average information in different domains. Specifically, are the limits on averaging stimuli over space the same as those for averaging in time (e.g., when stimuli are presented sequentially), and does this depend on the type of stimulus used?

Perceptual averaging

Although observers are able to estimate average properties from ensembles, they do not behave as

Citation: Florey, J., Dakin, S., & Mareschal, I. (2017). Comparing averaging limits for social cues over space and time. *Journal of Vision*, *17*(9):17, 1–13, doi:10.1167/17.9.17.

doi: 10.1167/17.9.17 Received October 11, 2016; published August 24, 2017 ISSN 1534-7362 Copyright 2017 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Downloaded from jov.arvojournals.org on 04/23/2021

though they are using all of the elements available. Dakin (2001) have demonstrated that when averaging the orientation of ensembles of Gabor patches, participants performed as if they were using only a subset of the total items in the array.[1] Similar results have been found for averaging of other low-level properties such as motion (Dakin et al., 2005) and size (Solomon, Morgan, & Chubb, 2011). Recently, Manning, Dakin, Tibber, and Pellicano (2014) found that when asked to determine the direction of motion of an array of 100 moving dots, children sometimes perform as if they are basing their judgment on only a single sample (dot).

This same subsampling effect has been found for face stimuli (Florey et al., 2016). When briefly presented with arrays of either faces with different gaze deviations or heads rotated in different directions, observers were able to judge the average of the set, although their estimate was based on their (effective) use of a subset of items. They were particularly limited when averaging the direction of gaze from a group, in some cases effectively basing their responses on a single face. We created classification images by correlating observers' responses with the distribution of locations and gaze or head offsets presented, which maps the stimulus locations that contributed to observers' judgements. This revealed that participants were biased toward using elements in the center of the array, an effect that was more pronounced for gaze direction than head direction.

There is conflicting evidence about the extent to which face-averaging tasks can be achieved in the periphery. Peripheral vision is limited by both reduced spatial resolution (Duncan & Boynton, 2003; Rovamo, Virsu, Laurinen, & Hyvärinen, 1982) and crowding, where elements presented in the visual periphery cannot be distinguished individually but rather appear cluttered (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). These limitations should increase the noise with which peripheral elements are processed, as has recently been shown for gaze direction in the periphery (Florey, Dakin, Clifford, & Mareschal, 2015; Loomis, Kelly, Pusch, Bailenson, & Beall, 2008; Palanica & Itier, 2015). This likely explains the limited contribution of elements falling near the edge of displays in a previous gaze-averaging study (Florey et al., 2016). Other studies however, report good averaging of facial properties when faces are presented in spatial arrays with elements falling in the periphery (e.g., Haberman & Whitney, 2009: some elements $>10°$ peripheral; Sweeny & Whitney, 2014: 3° peripheral); and Wolfe, Kosovicheva, Leib, Wood, and Whitney (2015) recently found that face expression was averaged equally well when faces around fixation were removed as when central faces were present.

Spatial limits

In tasks that are less limited by peripheral resolution than gaze perception, spatial integration still suffers from subsampling, suggesting that there are other limits on spatial integration. One possible limit is the spread of an observer's attention. Chong and Treisman (2005a) found that a dual task that encouraged a spread of attention improved participants' size averaging. Although this seems to suggest that a wider spread of attention improves performance on an averaging task, increasing the duration of the stimulus presentation does not. In another study, Chong and Treisman (2003) also found that reducing exposure duration from 1000 ms to 50 ms had little impact on size averaging. Sweeny and Whitney (2014) found that reducing the presentation time of a set of four gaze-direction stimuli from 1000 ms to 200 ms appeared to actually *increase* the number of elements being integrated. It seems, then, that there are global limits on integration within spatial ensembles (e.g., distribution of spatial attention and presentation time) that are distinct from limits on the processing of the individual elements within spatially distributed arrays (e.g., limited peripheral perception).

Earlier averaging studies indicate that peripheral presentation consistently leads to elevation of internal noise and does not affect sampling. For example, there is no net effect of overall region size (for fixed-size elements) on either orientation averaging (Dakin, 2001) or motion averaging (Dakin et al., 2005), even when large stimuli push elements far into the periphery. More directed psychophysics and modeling indicate that motion processing is limited by local noise in the periphery (Mareschal, Bex, & Dakin, 2008). More generally, peripheral vision is limited by crowding (Levi, 2008, review)—not, for example, acuity—and this includes processing of faces (Martelli, Majaj, & Pelli, 2005). It is known that crowding of orientation averaging only elevates internal or local noise, while attentional diversion only reduces sampling (Dakin, Bex, Cass, & Watt, 2009). Thus, there is a consistent and considerable body of evidence indicating that visual averaging of peripheral stimuli (whether crowded or uncrowded) is limited by internal noise but operates with similar sampling efficiency.

Temporal limits

Similarly, averaging of visual cues over time is not perfectly efficient. Gorea et al. (2014) report that when judging the average size of a temporal sequence of circles, participants performed as if they were using up to four out of eight elements. In most sequential averaging tasks, stimuli appear at fixation (e.g.,

Downloaded from jov.arvojournals.org on 04/23/2021

Albrecht et al., 2012; Corbett & Oriet, 2011; Haberman et al., 2009; Leib et al., 2014; though Haberman et al. do include a sequential task with peripheral stimuli), reducing the limiting effects of either the perception of any one element in the array (e.g., due to eccentricity) or the spatial distribution of attention. However, there are unique factors that influence sequential averaging, notably biases toward favoring particular temporal positions within the sequence. Researchers have used regression analysis to show that observers are biased by primacy (increasing the weighting of items appearing early in the sequence) and recency (increasing the weighting of later items; Gorea et al., 2014; Hubert-Wallander & Boynton, 2015). Hubert-Wallander and Boynton (2015) examined these effects using different types of stimuli and found stimulus-specific differences in the bias, with face expression and size averaging producing recency effects but position averaging producing primacy effects.

There is some debate as to the reason for these biases. Primacy could result from serial dependencies, a perceptual effect where each element in a sequence is biased to appear more like the item preceding it (Fischer & Whitney, 2014). If each element is influenced by the previous one in the set, responses will be biased toward the early elements, leading to a primacy effect. Alternatively, primacy may result from observers adopting a strategy of ignoring later samples, potentially because they have a limited capacity for integration. An efficient strategy would be to stop adding more information to the average computation when the resource cost of including it outweighs the potential improvement in accuracy; such behavior has been observed in both human and nonhuman primate observers (Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012).

One possible explanation for recency is a limit of attention or memory resources, resulting in early information being ignored or forgotten. Another is adaptive gain control (Cheadle et al., 2014), where elements that are consistent with the expected mean of the sequence thus far are upweighted and those that are inconsistent are downweighted. This type of strategy has been shown to produce recency in both simulated and human data.

Mechanism of averaging

How is averaging performed? The neural mechanism is not well defined. It is clear that we can perform less-than-perfect averaging, but beyond that our understanding is limited. Allen, Hess, Mansouri, and Dakin (2003) have shown that orientation averaging does not automatically pool estimates from luminance and contrast-defined elements, suggesting distinct averaging mechanisms for both. Haberman, Brady, and Alvarez (2015) provided evidence against a single generic mechanism for averaging of any type of stimulus. In their study, individuals' performance on low-level averaging tasks (size, color, orientation) did not correlate with their performance on averaging high-level face stimuli. However, they do report correlations within groups of similar stimulus type (high/low), suggesting some commonality within stimulus type. Contrasting evidence for domain-agnostic averaging comes from Florey et al. (2016), who found that observers were equally good at integrating information within a stimulus group (faces) as between stimulus groups (faces and 3-D cones). Whether the same mechanisms are employed in spatial and temporal averaging is less well understood. There are necessarily differences in early visual processing for these two types of stimuli, but it seems plausible that the higher level mechanism that integrates multiple elements into a single summary statistic may operate across stimulus types.

Noise paradigms

One method for examining performance on averaging tasks is using an equivalent noise (EQN) procedure that estimates two limits on observers' performance: *internal noise* and *effective sample size*. This is based on the assumption that when observers average, they first estimate the feature of interest for individual items in the set (e.g., the orientation of each Gabor in an array) before averaging a sample of these estimates. Internal noise refers to the observers' uncertainty about a single feature. Effective sample size tells us how many samples an ideal observer would need to average (given the internal noise on each sample) to achieve the observer's level of performance. For a stimulus containing n elements, a perfect observer would have no internal noise and an effective sample size of n, perfectly averaging the features of all the items into an accurate representation of the mean. In reality, observers display some amount of uncertainty associated with processing of individual items and their averaging strategy effectively subsamples the pattern. Observers behave as though they are using only a subset of the ensemble to estimate its average. To estimate internal noise and effective sample size, EQN experiments measure observers' averaging in the presence of different levels of variability of the feature of interest. When the variance is low (e.g., Gabors with similar orientation), performance is limited by the internal noise; if observers accurately perceive the orientation of one element, they will give a correct response. When

Downloaded from jov.arvojournals.org on 04/23/2021

variance is high (e.g., Gabors with widely differing orientations), the precision of any one estimate becomes less important, because the variance in the feature will swamp the influence of internal noise on individual elements. In this situation, the number of elements averaged will determine the precision of the observer's response. Modeling performance as variance in the stimulus, using an ideal observer, allows recovery of both internal noise and effective sample size.

Using the EQN method, previous research into spatial averaging has found that observers effectively use only $\sqrt{n}$ elements for orientation averaging (Dakin, 2001), or even fewer in the case of children averaging motion direction (Manning et al., 2014) or adults averaging gaze and head direction (Florey et al., 2016). Similarly, Gorea et al. (2014) found that observers sample only a subset of elements from a set of circles in a sequential size-averaging task.

Solomon and colleagues have employed a model related to EQN for integration of orientation and size stimuli (Gorea et al., 2014; Solomon, 2010; Solomon et al., 2011). The key difference in their "Noisy, inefficient but otherwise ideal observer" model of cue integration is that they separate the internal-noise term into two separate sources of noise: one that acts before the entire summary is integrated, either on individual stimuli or on "local pools" of subsets of stimuli (early noise), and one that acts at the level of the ensemble code, before a decision is made. The EQN analysis used in this study makes no assumptions about the source of internal noise; internal noise could be the result of uncertainty in processing the low-level properties of the face stimuli (e.g., in area V1) or at a higher level where gaze- or head-direction processing occurs (reportedly areas in STS; Calder et al., 2007; Perrett et al., 1985).

The current study

We have previously suggested that head direction may be a useful cue in acquiring a gist percept of a group of faces (a crowd), and that a serial average of the gaze direction of individuals may provide a more precise, albeit slower, average. Here we measure observers' ability to average head direction and gaze direction both in temporal sequences and across spatial arrays, using EQN analysis. The presentation duration and size of the spatial and temporal arrays are matched to allow a comparison with an equal amount of processing time available for each. This means that observers can make multiple saccades in the spatial condition, creating more naturalistic viewing conditions for crowd perception (Florey et al., 2016). Under both of these spatial and temporal conditions, we would expect that gaze-direction averaging should be similar to head-direction averaging, because the observers will be able to fixate the faces separately, eliminating the limits on peripheral processing of gaze direction. Alternatively, observers may not employ an efficient saccade pattern when presented with spatial arrays of gaze stimuli (i.e., they do not saccade to a new face each time or saccade between faces without processing each foveally), and as a result, they may not improve relative to brief presentations (e.g., 300 ms used by Florey et al., 2016).

We compare performance across two types of presentation condition (in space and in time) to determine if individuals who are good averagers (i.e., have low internal noise and high effective sample sizes) in one domain are also good in the other domain. We use a correlation analysis to examine how each EQN parameter correlates across stimulus type and presentation condition.

If averaging gaze direction and head direction share a source of *local* noise—for example, neural noise in primary visual cortex (area V1)—this would be reflected in a correlation in the internal noise between the two presentation conditions for each stimulus type (e.g., Figure 1a). Similarly, if there is a shared source of noise between processing spatial arrays and sequences of faces, then observers' internal noise should correlate between presentation conditions for each stimulus type (e.g., Figure 1c). Alternatively, if there are independent sources of noise which affect only one presentation condition, such as limited peripheral perception of eccentric faces in the spatial arrays, then we would not expect presentation conditions to correlate.

These same correlations can be carried out on the results for effective sample size to see if there are shared *global* limits for the stimulus types. For example, if effective sample size correlated between head and gaze cues for temporal sequences (e.g., Figure 1b), this would suggest that there exists a shared limit on the number of elements that can be integrated over time, perhaps due to constraints on short-term memory. If effective sample size correlates between presentation conditions (e.g., Figure 1d), this would suggest a generic global integration limit set by the *individual* rather than the stimulus properties.

Methods

Participants

Ten observers (seven women, three men) participated in the experiment, including one author (JF). All observers had normal or corrected-to-normal

Downloaded from jov.arvojournals.org on 04/23/2021

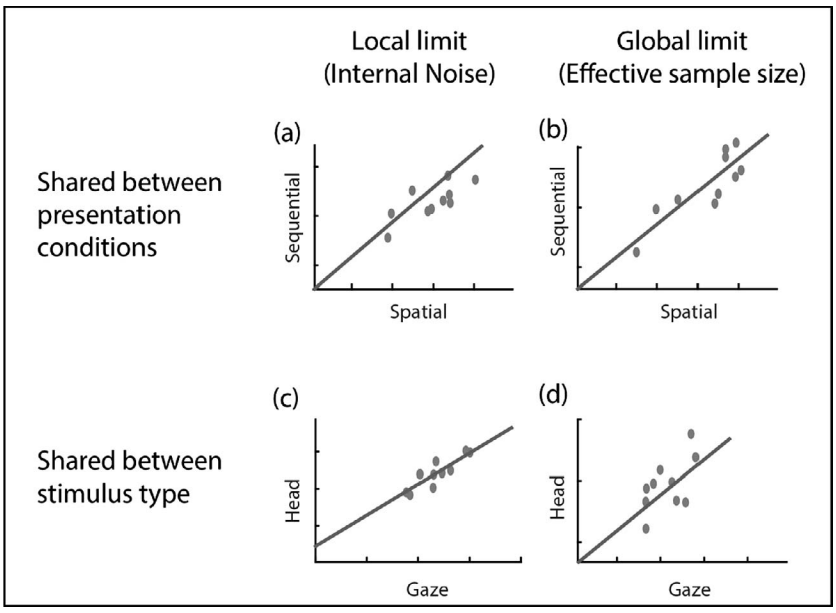

Figure 1. Hypothetical results to illustrate what different correlations reveal about shared limits. (A) If there are correlations between the internal-noise estimates for the two presentation conditions, this indicates a shared *local* limit on averaging. (B) As in (A) but for a shared global limit. (C–D) As for (A–B) but for limits shared between stimulus types rather than presentation conditions.

vision and gave informed consent according to the Declaration of Helsinki. All methods were approved by the ethics board at Queen Mary University of London.

EQN method

The two EQN parameters are estimated by measuring *observer noise* (specifically, observers' uncertainty on their estimate of the mean) as a function of changing *external noise*. The relationship between these data is described in Equation 1; observer noise is the sum of the internal and external sources of noise, divided by the number of samples used:

$$\sigma_{obs}^2 = \frac{\sigma_{int}^2 + \sigma_{ext}^2}{n_{samp}}, \quad (1)$$

where σ_{obs} is the observer's discrimination threshold, σ_{int} the internal noise, σ_{ext} the added external noise, and n_{samp} the effective number of samples used to estimate the mean.

In our experiment we quantified observer noise by estimating observers' threshold for discriminating whether a group of faces is looking on average to the left or right of direct. Thresholds were determined using a method of constant stimuli. Observers are presented with ensembles whose mean offset is either to the left or right of direct gaze and are required to indicate (reporting "left" or "right") the mean direction of gaze (or head direction) of the ensemble. By measuring performance repeatedly for a fixed number of offsets, we can fit a psychometric function to each observer's performance (proportion of trials identified as rightward) and estimate each observer's discrimination threshold for different levels of external noise. The standard deviation of the normal distribution from which the gaze head direction of each face is drawn corresponds to the external noise. At low external-noise levels (narrow standard deviation), the faces will all be looking in approximately the same direction, so observers are limited by how well they can estimate the direction of any individual face (internal noise). When the external noise is high the faces will be looking in dissimilar directions and the external noise will exceed the internal noise, so observers will now be limited by the number of samples they are able to average (Figure 2a). By measuring discrimination thresholds at a range of external-noise levels, we are able to fit a function to the data using Equation 1 to obtain estimates for each observer's internal noise and effective sample size, for each stimulus type and presentation condition (e.g., Figure 3).

Stimuli

Sets of eight gaze or head directions were generated for the spatial and sequential averaging conditions. The individual gaze-direction stimuli were generated by first randomly choosing a facial identity from a set of four

Downloaded from jov.arvojournals.org on 04/23/2021

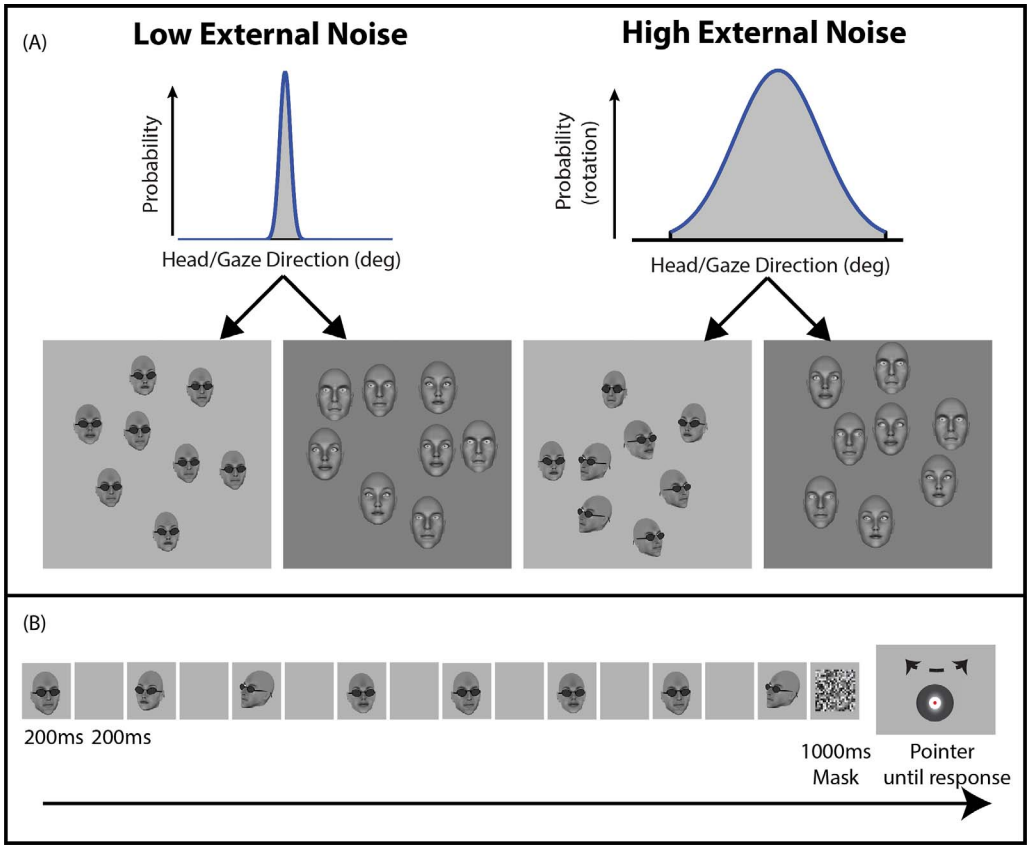

Figure 2. Examples of the stimuli presented in the four conditions. (A) Two normal distributions from which the direction of each gaze- or head-direction stimulus could be drawn. Below are examples of corresponding head- and gaze-direction stimuli that would be generated from these distributions, one with low external noise (all faces looking in the same direction) and one with high external noise (all faces looking in different directions). (B) A schematic depiction of a sequential head direction. Faces are presented in a sequence with 200-ms blank intervals between, followed by a noise mask and then a 3-D pointer to indicate the average direction.

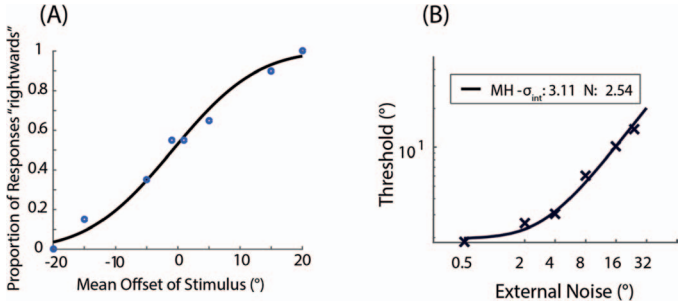

Figure 3. (A) The proportion of times the participant responded "rightward" to a given mean offset is plotted (blue circles) against the mean offset of the ensemble. A cumulative Gaussian function is fitted to these data (black line); its slope is the observer's discrimination threshold (which quantifies uncertainty about the ensemble mean). (B) An EQN plot. The thresholds (x) are plotted against the corresponding external-noise level. The black line is the fit from the model described in the text. The inset shows estimated internal noise (/sigma) and effective sample size (*N*).

synthetic faces (two female, two male) created using FaceGen software (Singular Inversions, Toronto, CA). The eyes were replaced with grayscale eye stimuli created in MATLAB (MathWorks, Natick, MA) to allow for precise manipulation of gaze offset. To create the individual head-direction stimuli, the same four synthetic faces were loaded into Poser (Smith Micro, Aliso Viejo, CA), a 3-D model-manipulation tool, and dark glasses were added to remove any cues from the gaze direction (Figure 2). Using Poser, we exported 1,800 frames of an animation of each head rotating between 90° leftward and 90° rightward, producing stimuli with steps of 0.1° of head direction. All faces were then scaled so that they would subtend 4° × 4° of visual angle during the experiment.

Individual face stimuli were combined to form spatial and sequential ensembles. For both the gaze- and head-direction stimuli, the offset of each face was drawn from a normal distribution. The mean of this distribution was determined by the offset value from the method of constant stimuli for the given trial, and

Downloaded from jov.arvojournals.org on 04/23/2021

the standard deviation was determined by the external-noise level being tested. In the spatial condition, faces were presented simultaneously with each face randomly positioned within a 12.5° radius from the center of the screen, such that no faces were overlapping (e.g., Figure 2a). For the sequential stimuli, faces were presented serially for 200 ms each, separated by 200 ms of a gray screen (e.g., Figure 2b). A small jitter (randomly chosen up to 1° in all directions) was applied to the position of each face in the sequence to avoid any apparent motion effects. A lighter background gray color was used for the head-direction stimuli, so that the edges of the faces were clearly defined.

Stimuli were presented on an Electron Blue CRT monitor (screen size: 30 × 40 cm) with a spatial resolution of 1,600 × 1,200 pixels operating at a frame rate of 85 Hz.

Procedure

Four sets of EQN parameter pairs were obtained for each of the four combinations of stimulus type and presentation condition (Spatial Gaze, Spatial Head, Sequential Gaze, and Sequential Head). For each EQN function, thresholds were obtained at six levels of external noise. The standard deviations of the normal distributions were 0.5°, 2°, 4°, 8°, 16°, and 24° degrees for gaze direction and 0.5°, 2°, 4°, 8°, 16°, and 32° for head direction. We set the highest noise level for the gaze-direction stimuli to 24°, to avoid generating stimuli that exceeded the physical limits of human gaze (i.e., gaze offsets >60°). Two blocks of 80 trials were collected for each external-noise level. Blocks included 10 repeats of the eight mean offsets in a random order, producing a total of 160 trials. The mean offset values presented within any block depended on the external-noise level of the block (to ensure even sampling of the psychometric function across conditions). For the gaze stimuli, noise levels below 5° standard deviation used offsets of −15°, −6°, −3°, −1°, 1°, 3°, 6°, and 15° from zero; and those above 5° used offsets of −20°, −10°, −5°, −1°, 1°, 5°, 10°, and 20°. For the head stimuli, three offset ranges were used: below 5° standard deviation: −6°, −2°, −1°, −0.5°, 0.5°, 1°, 2°, and 6°; for 8° standard deviation: −15°, −6°, −3°, −1°, 1°, 3°, 6°, and 15°; and for standard deviations above 8°: −30°, −10°, −5°, −1°, 1°, 5°, 10°, and 30°. Blocks for a single condition were collected in approximately hour-long sessions with a randomized order of external-noise levels.

Experimental control and stimulus presentation were controlled in MATLAB using Psychtoolbox (Brainard, 1997). In the spatial blocks the eight faces were presented simultaneously for 1600 ms. In the sequential blocks each face was presented for 200 ms separated by 200 ms of a blank screen followed by a 1000-ms noise mask. In both presentation conditions, the stimulus was followed by a 3-D response pointer which could be rotated with the mouse. The pointer consisted of a 3-D sphere with a red-and-white target drawn on the center. Moving the mouse rotated the sphere about its vertical axis to point in the direction the mouse was moved toward. The perceived gaze direction was taken as the orientation offset of this sphere when the observer clicked. The observer rotated the pointer and clicked to indicate when it was pointing in the mean gaze or head direction of the set of faces. No feedback was given, and the next trial commenced 200 ms following the response.

Threshold and EQN fitting

Observers' responses were converted to 1 (positive) or −1 (negative) to indicate an overall leftward or rightward response, respectively. Data from two separate runs for each participant were combined, giving 20 repeats at eight different offset levels. A cumulative Gaussian function was fitted to the proportion of times the participant responded rightward for each mean offset direction (Figure 3a) using a maximum-likelihood method. The standard deviation of this cumulative Gaussian function was taken as the discrimination threshold for the participant at a set level of external noise.

Discrimination thresholds quantify observer noise and (for a single participant and single stimulus/presentation combination) are plotted against the external-noise levels (Figure 3b). The EQN function (Equation 1) was then fitted to these threshold values (solid line, Figure 3b), yielding estimates of internal noise and effective sample size.

Results

The results for the EQN analysis are summarized in Figure 4a. A 2 × 2 (gaze/head, spatial/sequential) repeated-measures ANOVA was conducted for each of the two EQN parameters (internal noise and sampling efficiency). For internal noise, there was a main effect of stimulus type, $F(1, 9) = 22.1$, $p = 0.001$. Pairwise comparisons revealed that gaze-direction averaging was associated with significantly more internal noise than head-direction averaging ($p = 0.001$). There was no main effect of presentation condition, $F(1, 9) = 1.49$, $p = 0.25$, nor a significant interaction, $F(1, 9) = 2.19$, $p = 0.17$. The ANOVA for effective sample size revealed no significant main effect for stimulus type, $F(1, 9) = 0.588$, $p = 0.46$, or presentation condition, $F(1, 9) = 2.58$, $p =$

Downloaded from jov.arvojournals.org on 04/23/2021

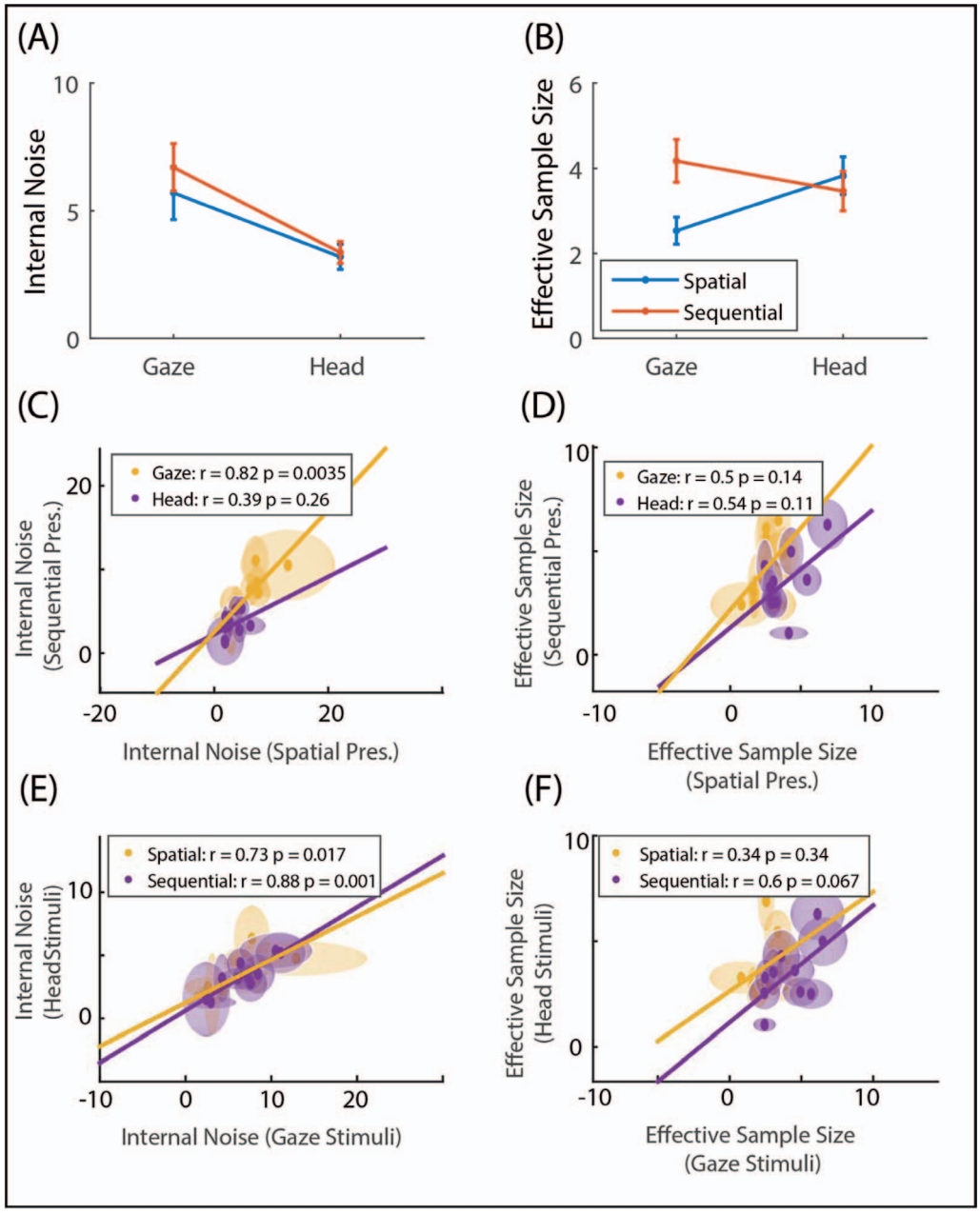

Figure 4. (A–B) Two plots showing the mean EQN parameters (internal noise and effective sample size) for the two stimulus types (x-axis) for each of the presentation conditions (red and blue lines). Error bars show ±1 standard error of the mean. (C–D) Scatterplots for the relationship between the two presentation conditions for each stimulus type (yellow = gaze, purple = head). Data from each individual are represented by the individual points, and a best-fit correlation line is drawn through the data. Figure legend shows the correlation coefficient *r* and significance *p* of the correlation. (E–F) As in (C–D) but for the relationship between the two stimulus types for each presentation condition.

0.12. There was, however, a significant interaction, $F(1, 9) = 20.4$, $p = 0.001$. Paired-sample t tests show that for the spatial presentation, gaze direction was associated with a significantly lower effective sample size ($M = 2.5$) than for head direction ($M = 3.8$), $t(9) = 2.9$, $p = 0.018$. For the sequential presentation, there was no significant difference between the two stimulus types, $t(9) = 1.6$, $p = 0.138$.

Taken together, these results show that observers are more uncertain about the direction of individual elements in gaze stimuli compared to head stimuli of the same size and presentation duration. The results for effective sample size suggest that observers can use fewer elements to average groups of spatially distributed gaze-direction stimuli than each of the other three combinations.

Downloaded from jov.arvojournals.org on 04/23/2021

Pearson's product–moment correlation analyses were performed for each EQN parameter, for the two stimulus types and the two presentation conditions (Figure 4C–F). For internal noise, a significant positive correlation ($r = 0.82$, $p = 0.003$) was found between the two presentation conditions for gaze stimuli but not for head direction. Across both presentation conditions, a significant positive correlation was found between the two stimulus types (spatial: $r = 0.73$, $p = 0.003$; sequential: $r = 0.88$, $p = 0.003$). For effective sample size, no significant correlations were found, though there was a borderline significant ($p = 0.067$) relationship between the two stimulus types in the sequential condition.

These results suggest that observers who have high internal noise for gaze stimuli will also have high internal noise for head stimuli for either presentation condition. Similarly, observers who have high internal noise for spatially distributed gaze stimuli also have high internal noise for sequentially presented gaze stimuli. The borderline significant correlation may suggest that observers who have high effective sample size for gaze will also have high effective sample size for heads, but only when both are presented sequentially.

Discussion

Using an EQN procedure we have compared averaging of head and gaze direction, for stimuli matched in size and presentation duration over both space and time. From these data, we estimated observers' internal noise (their uncertainty in estimating the direction of a single face from the array) and effective sample size (the number of samples they are able to effectively average) for ensembles of eight face stimuli. We report that head direction is averaged with a greater effective sample size than gaze direction in the spatial condition but not in the sequential condition. In both presentation conditions, gaze-direction judgements were associated with more internal noise than head direction. Correlation analysis revealed a relationship between observers' internal noise for the two stimulus types, and between the two presentation conditions for gaze stimuli. A possible relationship was also found between the effective sample sizes of the two stimulus types when sequentially presented, though this was only borderline significant.

The results for the internal-noise estimates are consistent with our previous results for gaze-direction averaging (Florey et al., 2016). This difference cannot be attributed to the peripheral presentation of the faces, since we report similar results using both spatial and sequential presentations, with the latter presenting all faces at the fovea. Although previous results have suggested similar precision in estimating gaze and head direction (Loomis et al., 2008), we find that for an averaging task, head-direction elements are processed with less uncertainty. This may be because attention is necessarily spread over either space or time, and gaze direction requires more focused attention to process with a high level of precision. Consistent with our previous findings, this suggests that in judging the direction of a crowd's attention, head direction is used for rapidly summarizing the direction of attention, and gaze can then be used to judge the interest of any individual within the crowd.

It is somewhat surprising that there was a significant difference in effective sample size between the two stimulus types for the spatial presentation. Although we had previously found this difference for briefly presented stimuli (300 ms), we expected that difference to disappear (or be largely reduced) here. This is because (a) observers would be able to make multiple saccades in the 1.6 s that the stimuli were presented for and (b) the stimuli were now larger, so that the limitations on peripherally processing gaze stimuli would be reduced. A possible explanation for our result is that observers may not accurately make saccades across the groups of faces (e.g., rather than foveate a single face, they saccade between two faces). This would mean that their average would still be limited by the fact that some faces fall in their periphery and so are not used in the average computation. As mentioned earlier, pervious research has shown that perceptual limitations as a result of eccentricity (acuity and crowding) have only affected internal noise and not sampling efficiency; however, it may be that the severity of the reduction of gaze perception in the periphery does cause the observed reduction in sampling. This difference between the two types of cue (head/gaze) is not present in the sequential presentation condition, suggesting that this is due not to the specific stimulus per se (i.e., that observers are poor at integrating gaze-direction signals) but rather to the spatial distribution of the elements in an array. This has important implications for averaging research that compares different types of stimuli, since the distribution of the elements as well as their peripheral visibility must be carefully controlled to avoid effects simply being the result of limited peripheral processing.

It is important to note a few caveats to these results. We report differences in effective sample size between spatial and temporal stimuli for gaze cues but not for head cues. Gorea et al. (2014) found that spatial and temporal effective sample sizes were not different, so our results for gaze direction are at odds with this result. This is most likely due to the type of stimuli

Downloaded from jov.arvojournals.org on 04/23/2021

used. Both this study and a previous study investigating gaze averaging (Florey et al., 2016) show that averaging of gaze over space is very poor, and although Sweeny and Whitney (2014) have found evidence of better gaze averaging, this may be a result of their use of stimuli that actually vary in head direction rather than gaze offset. In the present study, the presentation size and time were increased to try to improve the peripheral perception of gaze, but this did not produce the expected improvement in averaging performance. Gaze cues must always be contained within a surrounding face, so there will always be a limit on how large peripherally presented gaze cues can be, suggesting that gaze direction is not what people use when rapidly judging the average direction of a crowd's attention.

Although our stimuli are matched in presentation *time* and *size*, other possible factors may play a role in averaging. For example, our stimuli were spatially separated (maintaining a fixed density), but it is not clear how to equate a *spatial separation* with a *temporal* one, since a number of other factors will covary with this. For example, changing the temporal density by presenting more faces within a fixed time will decrease processing time of any individual face and increase the effects of backward masking (of one face on the previous one).

The correlation results for internal noise provide an interesting insight into what limits individuals in their averaging performance. We find that internal noise is highly correlated across the two types of stimuli when they are presented in a similar manner. This suggests that there is a source of noise that may be generic to these two different stimulus types. This source of noise most likely arises at an early level, potentially as a result of a common mechanism. In the case of head and gaze direction, both cues may at some stage be encoded as some positional offset—for gaze, of the iris within the sclera, and for heads, of the features (e.g., nose and glasses) within the face. If this shared mechanism is noisy, then this is a likely candidate for the shared source of internal noise.

The pattern of correlations between presentation conditions provides a less clear picture. Internal noise for gaze stimuli correlates between the two conditions but not for head direction. This result suggests that processing of a single gaze direction is independent of manner of presentation, but the same does not hold for processing of head direction. The difference observed for head direction could be explained by the fact that in the spatial condition, observers are required to spread their attention across multiple samples simultaneously, which may come at some cost to the precision with which each element is represented individually. The result for gaze is inconsistent with this interpretation; however, given the previously discussed limitations associated with spatial gaze stimuli and the low effective sample sizes observed here, it is possible that observers do not attempt to spread their attention in this condition and so do not suffer the costs to precision that they do when processing the spatially distributed heads.

The finding that effective sample size does not correlate between the two stimulus types in the spatial condition is potentially at odds with findings of Haberman et al. (2015), who report that individuals' averaging performance was correlated between two different face-based tasks (face emotion and identity). It may be that their correlation arose from a shared source of internal noise as opposed to similar sampling efficiency between tasks. Observers may have the same limits on their sampling efficiency between the two stimulus types, but because they are also limited by their peripheral perception of gaze, this correlation does not become apparent. The fact that the two stimulus types were weakly significantly correlated for sequential averaging but not for spatial averaging provides some support for this interpretation.

Our finding that there is no relationship in sampling efficiency between the two presentation conditions suggests that there is no generic limit on the integration of multiple samples independent of the way they are presented. Most likely, independent limits—such as spread of attention and sampling strategy in spatial ensembles, and short-term memory and temporal biases in sequential ensembles—have a greater influence on sampling efficiency (and performance) than any generic limit imposed by a single averaging mechanism.

Conclusions

In summary, we find that observers average head direction equally well over space and time. Gaze direction is also averaged well over time, though it is associated with higher internal noise. Averaging gaze over space was found to be severely limited compared to all other conditions, likely due to limitations in processing peripheral gaze stimuli. Gaze and head averaging have been shown to share a source of internal noise, likely occurring at the level of early visual processing. Global limits on averaging computations do not seem to be shared across presentation conditions, though there is some tentative evidence to suggest that there are shared global limits for temporal averaging.

Clearly there are many limits that must be considered with averaging; here we address some of the issues for simultaneously presented stimuli and suggest that

Downloaded from jov.arvojournals.org on 04/23/2021

care must be taken to ensure that any differences in peripheral perception are controlled for, even when using long presentation durations. In addition, when considering individual differences in averaging ability, it is important to consider whether performance is being limited by internal noise or sampling efficiency, as the two can vary independently.

Keywords: summary statistics, ensemble coding, gaze, social cues, averaging, equivalent noise, crowd attention

Acknowledgments

I. Mareschal was supported by a Leverhulme Trust grant RPG-2013-218.

Commercial relationships: none.
Corresponding author: Isabelle Mareschal.
Email: i.mareschal@qmul.ac.uk.
Address: Experimental Psychology, Queen Mary University of London, London, UK.

Footnote

[1] Note that estimates of effective sampling are based on noise paradigms which (a) quantify the effect of adding uncertainty on the attribute of interest on averaging and then (b) estimate the sample size that an ideal averaging system would require to achieve such a performance. While this sets a minimum sampling rate (i.e., we can use it to say the observer is averaging at least X samples), the exact number of samples may depend on the (likely nonideal) averaging strategy being used (so that other sources of noise may be limiting performance). We are thus careful to refer to *effective sample size*.

References

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, *74*(5), 810–815, doi:10.3758/s13414-012-0293-0.

Allen, H. A., Hess, R. F., Mansouri, B., & Dakin, S. C. (2003). Integration of first- and second-order orientation. *Journal of the Optical Society of America A*, *20*(6), 974–986, doi:10.1364/JOSAA.20.000974.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162, doi:10.1111/1467-9280.00327.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Calder, A. J., Beaver, J. D., Winston, J. S., Dolan, R. J., Jenkins, R., Eger, E., & Henson, R. N. A. (2007). Separate coding of different gaze directions in the superior temporal sulcus and inferior parietal lobule. *Current Biology*, *17*(1), 20–25, doi:10.1016/j.cub.2006.10.052.

Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., de Gardelle, V., Herce Castañón, S., & Summerfield, C. (2014). Adaptive gain control during human perceptual choice. *Neuron*, *81*(6), 1429–1441, doi:10.1016/j.neuron.2014.01.020.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404, doi:10.1016/S0042-6989(02)00596-5.

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13, doi:10.3758/BF03195009.

Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900, doi:10.1016/j.visres.2004.10.004.

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*, *138*(2), 289–301, doi:10.1016/j.actpsy.2011.08.002.

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A*, *18*(5), 1016–1026, doi:10.1364/JOSAA.18.001016.

Dakin, S. C., Bex, P. J., Cass, J. R., & Watt, R. J. (2009). Dissociable effects of attention and crowding on orientation averaging. *Journal of Vision*, *9*(11):28, 1–16, doi:10.1167/9.11.28. [PubMed] [Article]

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, doi:10.1016/j.visres.2005.07.037.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192, doi:10.1016/S0042-6989(97)00133-8.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision mak-

Downloaded from jov.arvojournals.org on 04/23/2021

ing. *The Journal of Neuroscience*, *32*(11), 3612–3628, doi:10.1523/JNEUROSCI.4010-11.2012.

Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, *38*(4), 659–671, doi:10.1016/S0896-6273(03)00265-4.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743, doi:10.1038/nn.3689.

Florey, J., Clifford, C. W. G., Dakin, S. C., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports, 6*, 32210.

Florey, J., Dakin, S. C., Clifford, C. W. G., & Mareschal, I. (2015). Peripheral processing of gaze. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 1084–1094, doi:10.1037/xhp0000068.

Gallup, A. C., Hale, J. J., Sumpter, D. J. T., Garnier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012). Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences, USA*, *109*(19), 7245–7250, doi:10.1073/pnas.1116141109.

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision*, *14*(9):22, 1–14, doi:10.1167/14.9.22. [PubMed] [Article]

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*(2), 432–446, doi:10.1037/xge0000053.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(11):1, 1–13, doi:10.1167/9.11.1. [PubMed] [Article]

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, doi:10.1037/a0013899.

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision*, *15*(4):5, 1–12, doi:10.1167/15.4.5. [PubMed] [Article]

Leib, A. Y., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception: A viewpoint-invariant mechanism to represent average crowd identity. *Journal of Vision*, *14*(8):26, 1–13, doi:10.1167/14.8.26. [PubMed] [Article]

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654.

Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., & Beall, A. C. (2008). Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, *37*(9), 1443–1457, doi:10.1068/p5896.

Manning, C., Dakin, S. C., Tibber, M. S., & Pellicano, E. (2014). Averaging, not internal noise, limits the development of coherent motion processing. *Developmental Cognitive Neuroscience*, *10*, 44–56, doi:10.1016/j.dcn.2014.07.004.

Mareschal, I., Bex, P. J., & Dakin, S. C. (2008). Local motion processing limits fine direction discrimination in the periphery. *Vision Research*, *48*(16), 1719–1725, doi:10.1016/j.visres.2008.05.003.

Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*, *5*(1):6, 58–70, doi:10.1167/5.1.6. [PubMed] [Article]

Palanica, A., & Itier, R. J. (2015). Effects of peripheral eccentricity and head orientation on gaze discrimination. *Visual Cognition*, *22*(9–10), 1216–1232, doi:10.1080/13506285.2014.990545.

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, doi:10.1038/89532.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *223*(1232), 293–317, doi:10.1098/rspb.1985.0003.

Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological Science*, *24*(8), 1389–1397, doi:10.1177/0956797612473759.

Rovamo, J., Virsu, V., Laurinen, P., & Hyvärinen, L. (1982). Resolution of gratings oriented along and across meridians in peripheral vision. *Investigative Ophthalmology & Visual Science*, *23*(5), 666–670. [PubMed] [Article]

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded

Downloaded from jov.arvojournals.org on 04/23/2021

arrays. *Journal of Vision*, *10*(14):19, 1–16, doi:10.1167/10.14.19. [PubMed] [Article]

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, *11*(12):13, 1–11, doi:10.1167/11.12.13. [PubMed] [Article]

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science*, *25*(10), 1903–1913, doi:10.1177/0956797614544510.

Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision*, *15*(4):11, 1–13, doi:10.1167/15.4.11. [PubMed] [Article]

Downloaded from jov.arvojournals.org on 04/23/2021