The contribution of foveal and peripheral visual information to ensemble representation of face race

Wonmo Jung
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea
Department of Science in Korean Medicine, Kyung Hee University, Seoul, South Korea

Isabelle Bülthoff
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea

Regine G. M. Armann
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea

The brain can only attend to a fraction of all the information that is entering the visual system at any given moment. One way of overcoming the so-called bottleneck of selective attention (e.g., J. M. Wolfe, Võ, Evans, & Greene, 2011) is to make use of redundant visual information and extract summarized statistical information of the whole visual scene. Such *ensemble representation* occurs for low-level features of textures or simple objects, but it has also been reported for complex high-level properties. While the visual system has, for example, been shown to compute summary representations of facial expression, gender, or identity, it is less clear whether perceptual input from all parts of the visual field contributes equally to the ensemble percept. Here we extend the line of ensemble-representation research into the realm of race and look at the possibility that ensemble perception relies on weighting visual information differently depending on its origin from either the fovea or the visual periphery. We find that observers can judge the mean race of a set of faces, similar to judgments of mean emotion from faces and ensemble representations in low-level domains of visual processing. We also find that while peripheral faces seem to be taken into account for the ensemble percept, far more weight is given to stimuli presented foveally than peripherally. Whether this precision weighting of information stems from differences in the accuracy with which the visual system processes information across the visual field or from statistical inferences about the world needs to be determined by further research.

Introduction

Ensemble representation is a concept specifying any mental representation achieved from combining multiple sensory measurements across space and/or time (see, e.g., Alvarez, 2011). It has been described and studied under different names, including global features, holistic features, statistical properties, summary statistics, and ensemble perception. The exact points stressed by each of these terms differ, but there is a fundamental concept at the basis of all of them: They all relate to redundant features or properties extracted from a scene consisting of multiple (similar) objects, thereby summarizing profuse visual information into a gist of a scene.

Besides studies investigating ensemble representation of low-level visual features of simple objects—such as size (e.g., Ariely, 2001; Chong, Joo, Emmanouil, & Treisman, 2008; Myczek & Simons, 2008), speed and direction of motion (Watamaniuk & Duchon, 1992; Watamaniuk, Sekuler, & Williams, 1989; Williams & Sekuler, 1984), or orientation (Dakin & Watt, 1997; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001)

Citation: Jung, W., Bülthoff, I., & Armann, R. G. M. (2017). The contribution of foveal and peripheral visual information to ensemble representation of face race. *Journal of Vision*, *17*(13):11, 1–12, doi:10.1167/17.13.11.

doi: 10.1167/17.13.11 Received November 7, 2016; published November 15, 2017 ISSN 1534-7362 Copyright 2017 The Authors

and location (Alvarez & Oliva, 2008; Morgan & Glennerster, 1991)—some studies have reported ensemble representation of complex high-level objects such as facial expression, gender, and identity (e.g., de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009, 2010). Haberman and Whitney (2007) for example, reported that observers could represent an average of facial expressions in a set from four to 16 faces presented for 2000 ms, while being unable to code or retain information about the emotional expression of individual set members.

Because of the long exposure time, however, this study could not fully reject the possibility that observers subsampled from the stimuli rather than averaging from the whole visual field. In a later study, Haberman and Whitney (2009) examined ensemble representation of facial expressions more systematically, comparing set sizes of four, eight, 12, and 16 faces and variable exposure durations of 50, 500, and 2000 ms. They reported consistent averaging performance of observers across various conditions that cannot be explained by an intentional subsampling strategy.

These and other studies (e.g., Haberman & Whitney, 2010) report ensemble representation of complex high-level visual information using large set sizes and thus suggest a rapid (as low as 50 ms), efficient (i.e., fairly independent of the number of items in a set), and flexible system. One aspect of ensemble representation that these studies do not consider is whether perceptual input from all parts of the visual field contributes *equally* to the ensemble percept.

In the human visual system, only the foveal region of the visual field—that is, the central 2° of visual angle—can be described with high fidelity. Information from other regions of the visual field, parafoveal and peripheral regions, is processed at lower resolution, suggesting that the quality and quantity of visual information transmitted from the fovea and more peripheral visual fields might differ. Visual stimuli located in different areas of the visual field might therefore contribute differently to ensemble representation. Im and Halberda (2013), for one example, reported different levels of internal noise in a size-discrimination task depending on whether the stimulus was projected onto foveal or peripheral areas of the visual field. Directly related to ensemble representations of faces, B. A. Wolfe, Kosovicheva, Leib, Wood, and Whitney (2015) demonstrated recently that accurate ensemble perception of the emotion of a crowd of faces does not *require* foveal input. When there *is* input from all parts of the visual scene, however, current data from Ji, Chen, and Fu (2014) suggest that foveal and extrafoveal input do play different roles when the visual system calculates ensemble representations from facial expressions.

Here, our goal is twofold: In the first place, we looked into the possibility that ensemble perception relies on weighting visual information originating from the fovea and from the visual periphery differently, rather than averaging equally across the whole stimulus set. To differentiate this from the subsampling already described, we call this strategy *fovea-biased* throughout this article, referring to an automatic, unconscious, differential weighting of incoming information by the visual system. At the same time, we extend the line of ensemble-representation research into the realm of race, another significant and widely studied feature of human faces. We examine whether averaging of race can occur, as has been reported for other face properties (identity: de Fockert & Wolfenstein, 2009; emotions: Haberman & Whitney, 2007, 2009, 2010; Ji et al., 2014; B. A. Wolfe et al., 2015; gender: Haberman & Whitney, 2007), and ask whether the visual system can extract the mean race of a set of faces when the time constraints and number of face stimuli presented at once do not allow conscious perception of single individuals.

In all experiments in this study, we used Asian and White face stimuli displaying the same average identity. More precisely, we averaged together individual identities within each race and created morphs between both averages. These average faces are unnaturally smooth and lack most idiosyncratic features that are used to identify individuals, so that race information is the dominating feature that changes between face stimuli. To distinguish between global and fovea-biased averaging, we compared observers' behavioral data to data obtained by modeling their expected responses following one of two hypothetical strategies: a global and a fovea-biased one. The former assumes that the race average estimated by observers is based equally on all faces in the set, while the latter assumes that the average is based mainly on the faces presented in the foveal part of the visual display (Experiment 1). By controlling the race composition of different parts of the visual display, we measured more precisely the size of the effect that foveal and extrafoveal regions might have on the perception of the mean race of the stimulus set (Experiment 2).

Methods

Experiment 1

In Experiment 1, we compared observers' behavioral data to modeled data assuming they used one of two potential strategies for averaging across all 12 faces in the stimulus set: For the global strategy, we assumed that the estimated race average is based on all presented faces equally, while for the fovea-biased strategy, we assumed that observers would give more weight to the faces in the center of a 3 × 4 face grid.

Participants

Fifteen White volunteers (seven women, eight men; age range: 20–35 years) residing in Tübingen, Germany, participated. Informed consent was obtained from all participants. All had normal or corrected-to-normal vision, were unaware of the purpose of the experiment, and were compensated for their time.

Stimuli

Using 32 Asian and 32 White male faces from the MPI face database (http://faces.kyb.tuebingen.mpg.de) and the Morphable Model algorithm (Blanz & Vetter, 1999), we created two prototype faces by averaging individual faces in both race groups. A set of 99 morphs was created by morphing gradually from one race prototype to the other, generating an Asian–White continuum of faces in 1% steps. We labeled the White prototype with a race level of 0% Asian, and the Asian prototype with 100% Asian; the 99 gradually morphed faces filled the levels 1% to 99% Asian.

All faces were presented in frontal orientation and in full color. We used a set of 12 faces in a 3 × 4 grid configuration (see Figure 1b), covering a visual angle of approximately 12° × 13° (following the stimulus presentation described in Haberman & Whitney, 2010, figure 1A).

We chose as test faces five faces evenly distributed on the Asian–White continuum (0%, 25%, 50%, 75%, and 100% Asian race level; see Figure 1a). While these test faces were clearly distinguishable from each other (it has been shown in a study with similar stimuli that a 25% distance between these race morphs is above observers' discrimination thresholds; see Bülthoff, Armann, Lee, & Bülthoff, 2015), they are unnaturally smooth and lack almost all facial information that can be considered idiosyncratic of a certain identity. The dominating features that allow observers to distinguish between these face stimuli are therefore related to their race difference. A new 3 × 4 face grid was created in each trial, by randomly selecting 12 faces from the five stimulus faces.

Task and procedure

The experiment was performed on a 22-in. computer screen using MATLAB 2009b (MathWorks, Natick, MA) and the Psychtoolbox extension (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). In each trial, a test set was presented for 250 ms following a fixation cross shown for 2000 ms. As in the study by Haberman and Whitney (2010) and other earlier studies on ensemble perception of faces already cited, participants were instructed not to focus on single faces but to just fixate in the middle where the cross had been shown and to distribute their attention evenly to all 12 faces. As literature on covert attention demonstrates, observers are able to disentangle their point of fixation from their point of attention allocation in this way (see, e.g., Carrasco, Williams, & Yeshurun, 2002; Pestilli & Carrasco, 2005). The very short presentation time, moreover, prevented our observers from making eye movements to scan the faces in each trial consciously. After a blank screen shown for 500 ms, a response face appeared in the middle of the screen. The response face (approximately 3.5° × 4° of visual angle) was taken from the Asian–White race continuum, chosen randomly in each trial. By moving the mouse, participants could scroll through all faces of this race continuum in an endless loop and choose one face with a mouse click (see Figure 1c). Participants were told to choose a face that corresponded to the average race of the faces seen

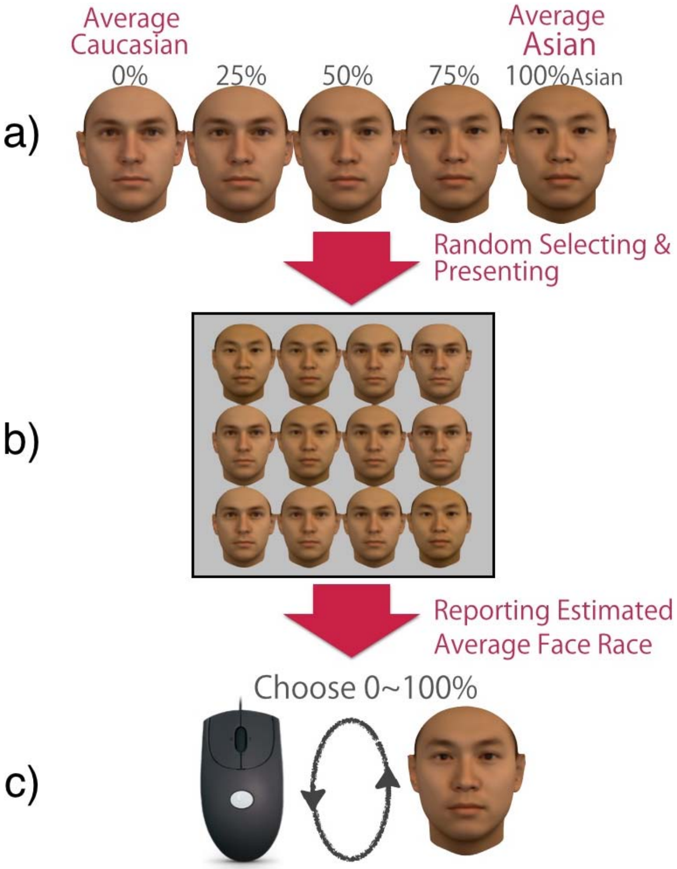

Figure 1. Stimulus selection and display. (a) Five face images from an Asian–White continuum were used as stimuli. (b) For each trial, a 3 × 4 grid (test set) was filled randomly with these faces. (c) Observers went through an endless loop of the entire race continuum and chose one face using a computer mouse.

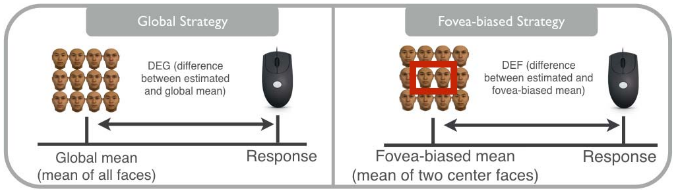
Global Strategy
DEG (difference between estimated and global mean)
Global mean (mean of all faces)
Response
Fovea-biased Strategy
DEF (difference between estimated and fovea-biased mean)
Fovea-biased mean (mean of two center faces)
Response

Figure 2. Analysis. For each test set of 12 faces, we calculated the global mean of all 12 faces (left) and the fovea-biased mean of only the two central faces (right), and compared observers' estimation to both hypothetical means. See text for more details and Figure 3 for a graphical representation of these comparisons.

in the preceding test set. A chin rest was used to maintain posture and a distance of 60 cm to the screen.

Based on earlier pilot experiments, the number of trials was fixed at 300 for each participant, so that an experimental session took approximately 60 min to complete. Before the task, participants were given detailed explanations about the concept of morphing faces between races and about what an average face represented, and had the opportunity to go through a few training trials.

Analysis

For each trial test set, we calculated two race means: the global mean, based on the Asian race level of all faces in the set (see Stimuli), and the foveal mean, based on the two central faces only. Since the faces in each set varied randomly between 0% and 100% Asian, we ran a Shapiro–Wilk test on the global mean data ($W = 0.99$, $p = 0.23$) and the foveal means ($W = 0.99$, $p = 0.08$) of the same stimulus selection, to make sure that our stimulus sampling was normally distributed.

For each set, we compared both means to the perceptual *estimations* given by each participant to deduce whether participants were following a global or a fovea-biased strategy (see Figure 2). To this end, based on the same estimations from our observers, we calculated the difference between estimated and global mean (DEG) and between estimated and foveal mean (DEF). Since DEG and DEF scores are the result of applying different calculations to the same data, their distributions have different dispersion values, which do not allow for direct statistical comparison. We therefore compared each distribution to hypothetical random distributions (as outlined later) to determine which strategy was more probable—that is, further away from random behavior.

If participants followed a random response strategy and randomly selected a mean race value (i.e., their answers varied between 0 and 100% independently of the test sets), the DEG and DEF values would also be normally distributed. If, however, participants estimated the average race of the set based on all faces (global mean) or on the most central faces (foveal mean), DEG or DEF values would be closer to 0 (i.e., closer to the global or foveal mean, respectively) than values for random data, and their distributions would therefore be narrower than in the random case.

We generated random response distributions simulating observers randomly selecting their responses using the Monte Carlo method (10,000 iterations), and compared the distributions of random DEG and DEF values to those calculated from observers' data (Figure 3). To compare the narrowness of the distributions

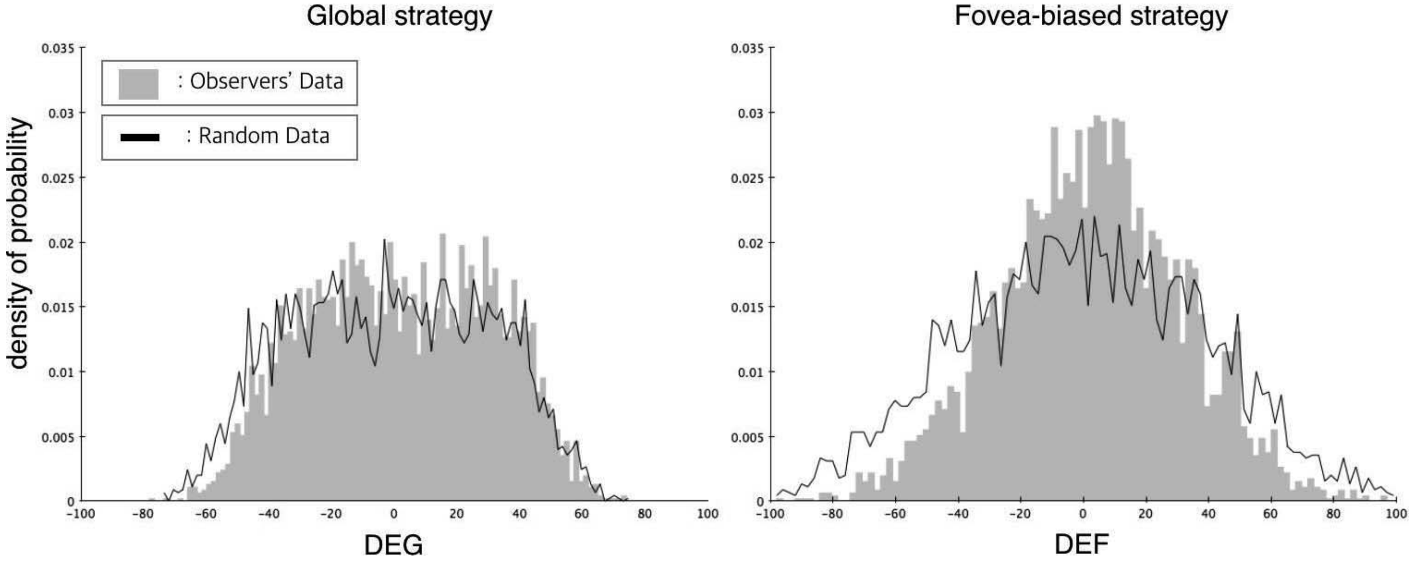

Figure 3. Probability distributions. Observers' distributions of their difference between estimated and global mean (DEG) and their difference between estimated and foveal mean (DEF) (left: global; right: fovea-biased) were compared to randomly generated DEG and DEF distributions. For more details, see text.

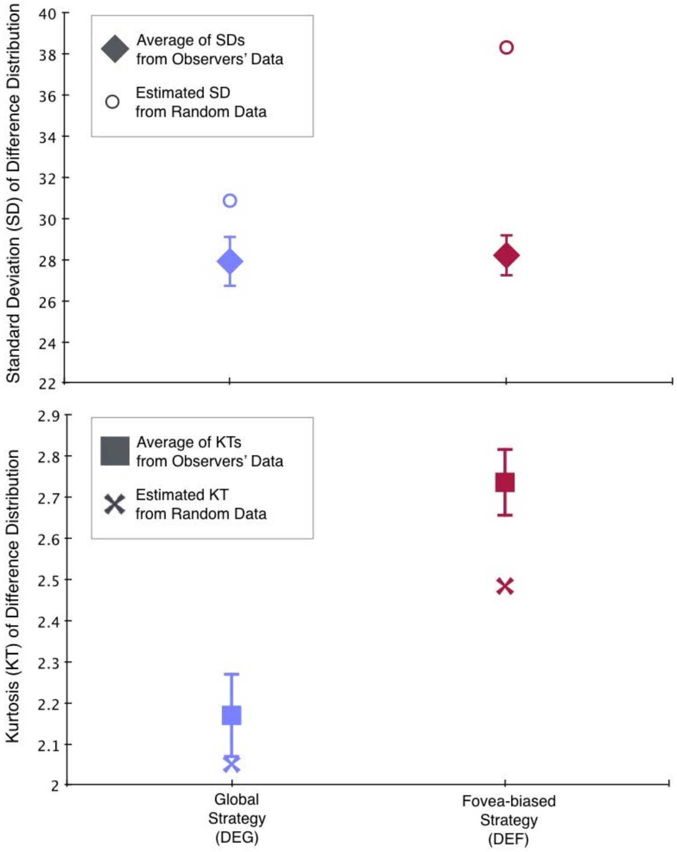

Figure 4. Standard deviations (top) and kurtosis values (bottom) of global (DEG) and fovea-biased (DEF) distributions calculated across observers and from random data.

quantitatively, we compared standard deviations (SDs) and kurtosis (KT) of the DEG and DEF distributions obtained from random and observers' data (see Figure 4). One-sample t tests (two-tailed) were used to judge whether the SD and KT of observers' data distributions differed significantly from the SD and KT generated from random data for both strategies.

Results

Distributions of DEG and DEF values obtained from all participants, as well as distributions derived from random data, are shown in Figure 3. A cursory look at the graphs reveals that participants' DEF distribution is narrower and more peaked than their DEG distribution, and, more importantly, that participants' DEF distribution appears narrower and more peaked than their corresponding randomly generated DEF responses. There is no visible difference in the shape of observers' and corresponding random DEG distributions. SDs and KT of DEG and DEF distributions calculated across all participants are shown separately for the global and fovea-biased strategies in Figure 4. If we look at just observers' data in both graphs in Figure 4, it is evident that the global (DEG) and fovea-biased (DEF) distributions have almost identical SDs, while the corresponding KT values assume a platykurtic shape for DEG and a narrower, approaching-normal distribution for DEF values. One-sample t tests show lower SDs ($M = 27.9$, $SEM = 1.19$) for DEG distributions from observers' data than for corresponding random data ($M = 30.9$, $t = -2.49$, $df = 14$, $p = 0.025$, $d = 0.64$); the difference, however, is much larger for DEF distributions (observers' data: $M = 28.2$, $SEM = 0.97$; random data: $M = 38.3$, $t = -10.4$, $df = 14$, $p < 0.001$, $d = 2.7$). For the assumed global strategy (DEG), there is no significant difference between KT values of observers' distributions ($M = 2.17$, $SEM = 0.10$) and of simulated random behavior ($M = 2.05$, $t = 1.13$, $df = 14$, $p = 0.278$, $d = 0.29$). For the assumed fovea-biased strategy (DEF), however, we find significantly higher KT values for observers' distributions ($M = 2.74$, $SEM = 0.08$) than for simulated random behavior ($M = 2.48$, $t = 3.14$, $df = 14$, $p = 0.007$, $d = 0.81$).

Discussion

In sum, as shown by SDs for estimated and random distributions, our observers showed—for both strategies—better estimation of the race average than random data would predict; the difference was, however, far greater for the assumed fovea-biased (DEF) strategy than for a global (DEG) strategy (see Figure 4). KT values did not show a difference between participants' and random data for the global (DEG) strategy, while for the fovea-biased (DEF) strategy they were higher for human observers than for random data. These (indirect) comparisons and the distributions in Figure 3 suggest that the estimated race averages obtained from our observers were closer to a fovea-biased mean based on the two central faces than to a global mean based equally on all 12 faces of each set.

Since our participants were not given enough time to scan individual faces (note that the duration of fixations to faces is usually well above 300 ms; see, e.g., Armann & Bülthoff, 2009; Henderson, Williams, & Falk, 2005), and were instructed to fixate in the middle of the stimulus set while trying to get an overall impression of the whole display, what we call here a fovea-biased strategy is very probably not related to conscious allocation of attention. As mentioned in the Introduction, ensemble representation for facial expression has been shown to be equally accurate with or without foveal input (B. A. Wolfe et al., 2015), which suggests a covert automatic process, even if this process is attention driven. Because we found in our experiment that observers' race estimation of the stimulus set is nevertheless closer to a subset average of the two foveally presented faces than to the average across the whole set, the question arises whether different parts of the visual display are weighted differently. To systematically estimate the influence from foveally and peripherally presented faces, in Experiment 2 we manipulated and controlled the race information displayed across faces in both parts of the visual display.

Experiment 2

Our findings in Experiment 1 suggest that observers base their race estimations more on the two central faces of each face set rather than on a global average calculated over all faces equally. The design of Experiment 1 was based on contrasting these two slightly extreme strategies (all faces equally versus only two center faces) and did not allow us to directly assess the influence of central and more peripheral faces in estimating face race. Experiment 2 was done to measure and more precisely compare their respective roles using a slight variation of the same paradigm. Here, instead of randomly choosing the average race of the whole set in each trial, we controlled how Asian or White (in race level) the center two faces were versus the 10 peripheral faces. Comparing actual race levels in the center and the periphery to observers' estimation for each set allowed us to derive a measure for the respective influence of both parts of the visual display.

Participants

Another 23 White volunteers (10 women, 13 men; age range: 21–43 years) residing in Tübingen, Germany, participated in Experiment 2. Informed consent was obtained from all participants. All had normal or corrected-to-normal vision, were unaware of the purpose of the experiment, and were compensated for their time.

Stimuli

In Experiment 1 we randomly selected faces for each set, which resulted in a large variety of global race-average values. To compare the effect of central and peripheral faces more precisely in Experiment 2, we modified the choice of faces in two ways to create stimulus sets. First, the 12 faces in each grid were separated into two groups—two central faces and 10 peripheral faces (compare to Figure 2). Second, we used fixed race averages for each group, which could take only one out of three values (25%, 50%, or 75% Asian race level). The selection of face morphs for obtaining these average values was done randomly in each trial.

Task and procedure

Task and procedure were identical to Experiment 1, with the following exceptions. In each trial, we randomly selected a central and a peripheral combination of faces to obtain one of the three fixed means. With the combination of three central and three peripheral race levels, our design consisted of nine conditions with 40 trials per condition, yielding a total of 360 trials. Three of the nine conditions were homogeneous conditions, showing the same average values in central and peripheral faces (central/peripheral: 25%/25%, 50%/50%, 75%/75%). In all other combinations, central and peripheral faces had different race averages.

Results

The three homogeneous conditions served as control conditions, allowing us to control for observers' potential bias in judging the mean race of the display even when all faces in a grid have the same race level (see Figure 5). For the 25% and 50% conditions, participants estimated the average to be more Asian than it was, while in the 75% condition they estimated it to be less Asian. As a result, estimated averages differed less from each other in those three conditions than real average values. We cannot determine at this point whether these biases stem from less precise visual information processing in the periphery or from a postperceptual race bias affecting the whole visual display. In order to examine what effect incongruent race information in the center and periphery has on observers' perception of the whole display, we analyzed

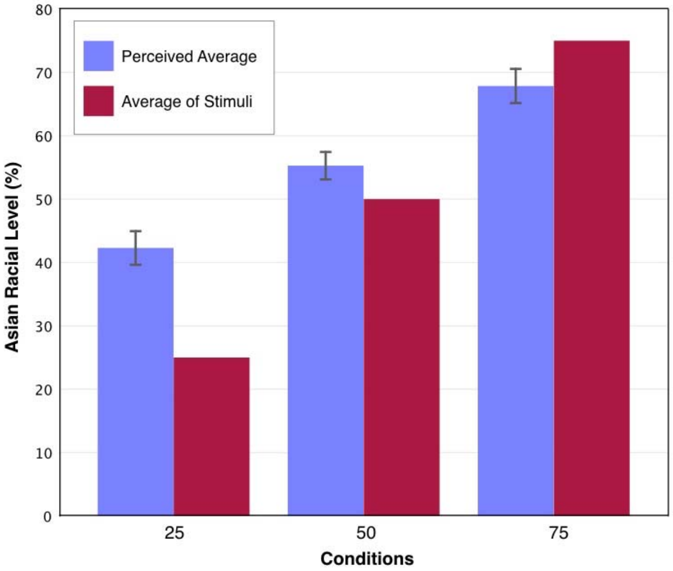

Figure 5. Estimated and real race averages for the homogeneous conditions. Error bars show the standard error of the mean. For more details, see text.

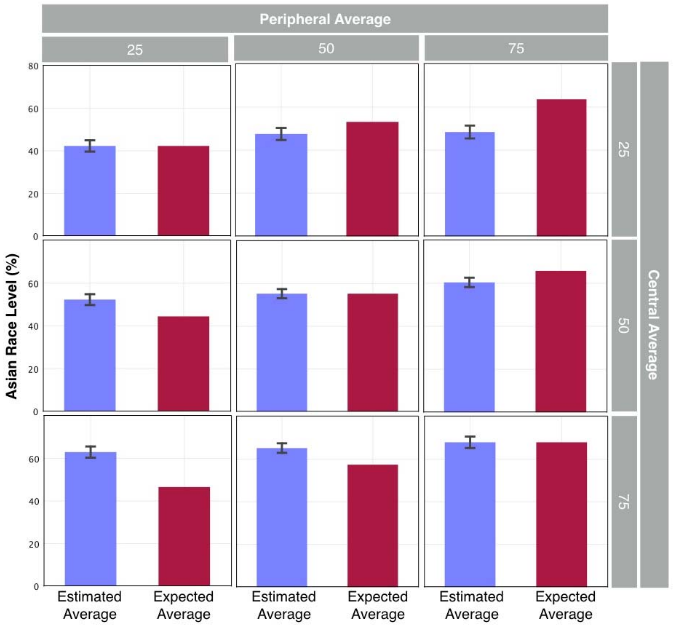

Figure 6. Expected (red) and estimated (blue) race averages for all conditions. Expected averages were calculated using observers' estimations in the homogeneous conditions (25/25, 50/50, 75/75) as a baseline (see Figure 5). Estimated values are observers' responses; they are by definition identical to the expected values in the homogeneous conditions. Error bars show the standard error of the mean. For more details, see text.

their estimations relative to their own percept of the homogeneous conditions. Therefore, for both strategies, we used observers' estimated race means in the homogeneous conditions as a race baseline to calculate *expected* averages in all nonhomogeneous conditions (see Figure 6) in the following way.

For the global strategy, by definition, each face contributes equally to the *expected* average of a set (global mean). The mean race values for the central and peripheral regions were thus exchanged for the race baseline values, according to the data from the homogeneous conditions, to calculate an expected average. For example, for the condition consisting of a 25% peripheral average and a 50% central average, the expected average corresponds to (10 × race baseline [25] + 2 × race baseline [50])])/12.

We calculated expected averages for all conditions, then compared those to observers' actual race estimations for each condition (see Figure 6). A cursory look at the figure shows that calculated expected averages vary across conditions in larger degree according to changes of the peripheral average (along the horizontal axis) than to changes of the central average (along the vertical axis). This is not surprising, because the peripheral group contains more faces than the central group. However, race estimations obtained from our participants show the opposite pattern: Their estimations are more affected by changes of the central average (vertical axis) than by changes of the peripheral average (horizontal axis). A two-way ANOVA on observers' race-average estimations with the factors center race level (25, 50, 75) and periphery race level (25, 50, 75) was performed to examine the influence of average race levels of center and peripheral regions on the reported average race level. Note that in this design, the two factors to be combined orthogonally are not location versus level, but the average race levels of each area (fixed at either 25%, 50%, or 75%). The test revealed a significant main effect for periphery race level, $F(2, 198) = 4.66$, $p = 0.011$, $\eta^2 = 0.032$, and for center race level, $F(2, 198) = 42.5$, $p < 0.001$, $\eta^2 = 0.29$, but no significant interaction between the two, $F(4, 198) = 0.24$, $p = 0.913$, $\eta^2 = 0.003$.

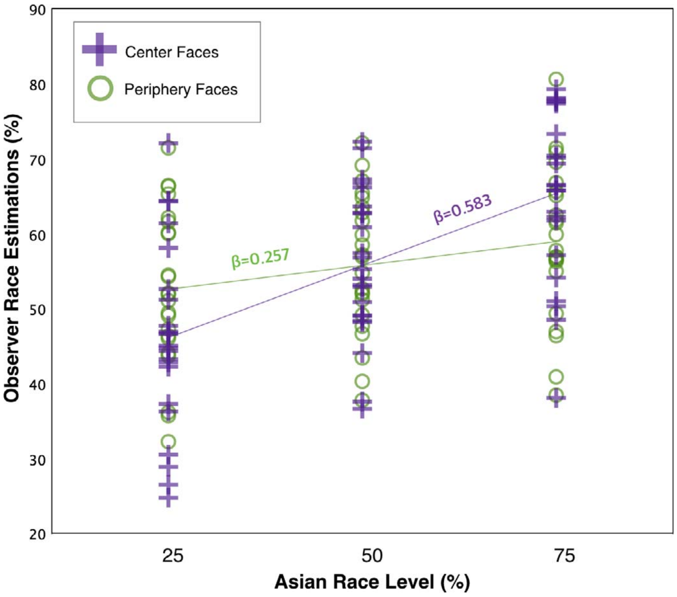

Figure 7. Simple linear regressions on participants' estimations for the three race levels (25%, 50%, 75% Asian) averaged across either central (cross) or peripheral (circle) faces only. For more details, see text.

While the ANOVA showed significant main effects for both center race level and periphery race level, we found their effect sizes to be very different. To get a closer look at how center and periphery faces might influence observers' race estimation of the test set differently, we again assumed two distinct strategies, a global one and a foveal one, and subdivided each observers' race-estimation data twice: once by race level of center faces (25, 50, 75) and a second time by race level of periphery faces (25, 50, 75). We then applied linear regression to each data set (note that both sets contain the same data) and compared the slopes of the resulting regression lines. As an illustration (see Figure 7), the "25% Asian Race Level" bin for the center-faces strategy consists of all trials with a 25% center race level (120 trials), independent of the race average in the periphery (i.e., containing all potential race levels at the peripheral location). The same bin for the periphery-aces strategy consists of all trials with a 25% periphery race level, independent of the race average of the center faces. The same grid consisting of two 25% Asian faces and 10 75% Asian faces was thus labeled 25% Asian in the center-faces condition but 75% Asian in the periphery-faces condition.

The regression analyses show that both central and peripheral faces influence observers' race estimation significantly, with a standardized beta coefficient of 0.257 for periphery faces, $F(1, 67) = 4.75$, $p = 0.033$, $R^2 = 0.066$, and 0.58 for center faces, $F(1, 67) = 34.4$, $p < 0.001$, $R^2 = 0.34$. Using linear hypothesis testing, we found that there is a significant difference between the two regression coefficients, $F(1, 204) = 19.53$, $p < 0.001$. In line with the different effect sizes we find in the ANOVA, these results illustrate that while the peripheral race level is averaged over 10 faces and the central race level over only two, the information from those two center faces yields a regression beta coefficient almost three times as large as the one for the 10 peripheral faces.

Discussion

In Experiment 1 we tested whether race averaging from a set of faces is possible in the same way that averaging of face expressions, gender, and identity has been reported in earlier studies. We also tested whether observers' race estimations are closer to a global average that takes all faces of the test set *equally* into account or to an average of only the two central faces closest to the point of fixation. The design of Experiment 1 did not, however, allow for a more fine-grained assessment of the potentially weighted influence of each part of the visual display. In Experiment 2, we therefore tested more systematically whether observers' race estimations are more influenced by the physical race makeup of the center or the periphery of the visual display. We find that the two central faces have a stronger influence on observers' ensemble percept than the 10 faces in the periphery.

General discussion

In sum, our experiments show that the visual system can estimate the average race of a crowd even when time constraints and the number of faces presented at once do not allow conscious perception of individuals, but that it gives far more weight to the faces perceived foveally than to faces in the periphery. This suggests that our first gist of a crowd of people would be strongly biased toward the faces we directly look at. The short presentation duration and high number of stimuli in the current and other studies on ensemble perception suggest that this is due not to active subsampling of visual information or a cognitive strategy but rather to an unconscious covert process, as has been proposed recently by B. A. Wolfe et al. (2015).

B. A. Wolfe et al. also report that occluding foveal information when presenting sets of faces in an ensemble-representation task does not impair observers' ability to estimate the average emotion (in their case) of the whole set—a finding that at first glance seems to contradict our results. One could assume that face race might be processed differently from emotion, and since the current study is the first one to systematically extend the line of ensemble-representation research from identity, emotion, and gender to face race (but see Thornton, Srismith, Oxner, & Hayward, 2014), it does not provide any argument for or against that assumption. While our morphing technique—that is, first creating two central tendencies by averaging within an Asian and a White face set, then defining the differences between those two averages as the predominant differences between the two races—does yield stimuli whose main dominating feature is race, it is of course not possible to completely disentangle identity and race information in faces. Unlike in experiments on ensemble percepts of identity and emotions (but in fact similar to ensemble experiments on gender), race morphs, as soon as the difference between two faces becomes too large (e.g., 50% or 75% in our continuum; see Figure 1), are perceived as different people. One might thus wonder whether some discrepancies between our results and earlier findings might have to be attributed to this technical issue, and further research using race morphs while exploring aspects of ensemble perception that have been investigated for identity—such as viewpoint-invariant representations (Leib et al., 2014) or multiple levels of ensemble representations (Haberman, Brady, & Alvarez, 2015)—might help to elucidate this question. Regarding the findings presented here, however, another recent study (Ji et al., 2014) using a very similar design to ours has explored how faces in foveal and extrafoveal vision contribute to an ensemble percept of emotion (i.e., a feature, unlike race, that can be manipulated within, and thus independently of, identity), and came to the same conclusion: Faces in foveal vision are given more weight than faces in the periphery. These similar findings from a face domain that has been studied before in ensemble representation make it unlikely that there is a different mechanism behind race averaging. Some work on ensemble representations of low-level features has also found that when observers' attention is drawn to a particular item in a set, mean estimation is biased toward the specific characteristic of that item (e.g., de Fockert & Marchant, 2008).

Rather than providing conflicting results, we would argue, the current study and the one by Ji et al. on one side and B. A. Wolfe et al. on the other side are looking into different aspects of ensemble perception of high-level face features. B. A. Wolfe et al. demonstrate that observers *can* judge the mean emotion of a set of faces in the presence or absence of foveal visual information (but note that feedback was provided after every trial). This is, however, no indication that the visual system does *not* rely more on foveal information when it is available. Observers in the occluded and nonoccluded conditions in those experiments might have pragmatically adjusted their strategy to reach the highest possible performance on the task at hand. When encountering a crowd of people in real life, however, the visual system's natural approach might be different. Furthermore, our experiments show how stimuli outside of the fovea are indeed taken into account in extracting an ensemble percept (even though the effect of the periphery in our experiments is surprisingly weak), which is in accordance with the results of B. A. Wolfe et al. in particular, and of other studies on ensemble perception.

There are a lot of examples in face research where strategies in a task are adapted to what is available. Successful face recognition from highly distorted, blurred, or selectively occluded images (see, e.g., Schyns, Bonnar, & Gosselin, 2002; Sinha, Balas, Ostrovsky, & Russell, 2006) demonstrates how flexible face-perception processes are in general. The most obvious example is probably prosopagnosic observers, who often excel at tasks that are used to diagnose low face-recognition skills by employing all kinds of compensatory strategies. These same prosopagnosic observers nevertheless do not have typical face-processing abilities, and struggle in social interactions in their daily lives (e.g., Esins, Schultz, Stemper, Kennerknecht, & Bülthoff, 2016).

What we suggest is thus that when we look at natural everyday crowd situations, rather than a trained feedback task in the lab, observers do make eye movements and bias their mean estimation toward information processed foveally—but they also integrate some information from across the whole visual field into their percept, and they do so in an automatic, covert process.

Objects outside the focus of attention are perceived less clearly and with lower contrast (see, e.g., Carrasco et al., 2002; Pestilli & Carrasco, 2005). Given differences in cell types, density, and connections to higher processing levels across the retina, objects outside the foveal field of view are also processed at lower resolution (e.g., Dacey & Petersen, 1992; Livingstone & Hubel, 1987; Schiller, Logothetis, & Charles, 1990). As Alvarez (2011) has illustrated, it makes intuitive sense for the visual system to give more weight to reliable than to unreliable information, and computer simulations show that error distributions are lower for such a precision-weighted average than for an average that weighs all information equally.

An alternative way of thinking about these results, proposed by Haberman et al. (2015), is that they could reflect the presence of two different representations—that is, an exemplar representation for the limited number of directly attended central faces and an ensemble representation including peripheral information. Since the processes underlying exemplar and ensemble representations are not identical, but might be mutually interactive, one could hypothesize that participants' average race estimation is based on some combination of both representations.

A general question about ensemble perception of faces that arises from the current research is whether the brain assumes a certain degree of reliability for weighting faces in different parts of the visual display, and whether this degree of reliability is drawn from statistical inferences (knowledge, experience) about the world. Alternatively, the visual system might simply use a heuristic to calculate averages, based for instance on visual resolution in the retina declining with eccentricity. One way of investigating this aspect further would be to present all face stimuli in a set foveally but sequentially across time. When eliminating visual-field differences in this way, Haberman and Whitney (2009) showed that ensemble information about facial expression is integrated over time and robust even for large set sizes. Comparing the temporal averaging process in such a continuous stream with no differences in visual resolution to the visual averaging from the stimulus grid used here and in numerous earlier studies on ensemble perception could be a starting point for further investigating whether statistical inferences or heuristics are the basis of the weighted averaging we find.

Haberman and Whitney (2010) have also shown that emotional deviants are not taken into account when an average is calculated across a visual set of faces. Regarding the significance of race and racial outliers in everyday life, it would be interesting to test whether the same holds true for sets of faces manipulated for race. In fact, some of our results on ensemble perception (unpublished data) suggest that more weight is given to own-race faces in race estimation. While these are preliminary results awaiting confirmation by more thorough testing, other-race faces could be interpreted as outliers, thus supporting the findings of Haberman and Whitney (2010) for emotion and suggesting that they hold for face race too (but see Thornton et al., 2014). Manipulating face stimulus presentation in terms of both visual eccentricity and time (by sequential presentation) would allow further insight into inferential or heuristic strategies at the base of ensemble perception when it comes to outlier detection and processing beyond effects of local saliency.

Note that our assumption of a weighted averaging strategy (due either to differences in accuracy across the visual field or to real-life experience) is not the only possible explanation for our results. It could also be possible that the visual system does assign equal weights to all incoming information but that due to a loss in resolution and contrast, as well as crowding occurring in the periphery, peripheral input has a lower impact on the calculation of the average. To distinguish between such a purely system-inherent mechanism on the one hand and attention-driven or cognitive strategies on the other, one could for example estimate and integrate the peripheral loss of information into the model, or display peripheral stimuli in such a way that the loss is compensated (e.g., manipulating size, contrast, or spacing), and see if the central bias persists under these circumstances.

What we have tested here so far is an ecologically valid crowd situation where faces in the periphery seem of lesser resolution and more crowded, from the point of the observer; future research will have to show whether the limitations of the visual field are the main driving force behind the central bias we and others find in ensemble representations of faces.

Conclusion

To summarize, we show here that observers can average across multiple faces when asked to judge the mean race of a group, similar to judgments of mean emotion or gender from faces, and similar to ensemble representations in other, more low-level domains of visual processing. We further find that the averaging process appears to take information from the whole visual field into account, but that far more weight is given to those faces presented foveally than peripherally. Further research has to show whether this effect simply stems from inaccuracies in the processing of information from the periphery of the visual field or if it represents a way of precision-weighting information. If the latter is true, such a strategy could then be due to differences in the accuracy with which the visual system processes information across the visual field, or to statistical inferences the brain makes about the world and the faces we encounter.

Keywords: ensemble perception, *summary statistics*, *face recognition*, *face race*

Acknowledgments

We want to thank Ian M. Thornton for providing the initial idea for the research that went into this study, for fruitful discussions while the work was being done, and for helpful comments on the manuscript. Part of the work that went into this publication was supported by a scholarship by the DAAD (German Academic Exchange Service)—NRF (National Research Foundation of Korea) bilateral scientist exchange program awarded to Regine Armann.

Commercial relationships: none.

Corresponding author: Regine G. M. Armann.
Email: regine.armann@tuebingen.mpg.de.
Address: Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Armann, R., & Bülthoff, I. (2009). Gaze behavior in face comparison: The roles of sex, task, and symmetry. *Attention, Perception, & Psychophysics*, *71*(5), 1107–1126.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *SIGGRAPH '99 proceedings of the 26th annual conference on computer graphics and interactive techniques* (pp. 187–194). New York: ACM Press/Addison-Wesley.

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

Bülthoff, I., Armann, R. G. M., Lee, R. K., & Bülthoff, H. H. (2015). The other-race effect revisited: No Effect for faces varying in race only. In S.-W. Lee (Ed.), *Recent progress in brain and cognitive engineering* (pp. 153–165). Dordrecht, the Netherlands: Springer.

Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision*, *2*(6):4, 467–479, doi:10.1167/2.6.4. [PubMed] [Article]

Chong, S. C., Joo, S. J., Emmanouil, T. A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics*, *70*(7), 1327–1334.

Dacey, D. M., & Petersen, M. R. (1992). Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proceedings of the National Academy of Sciences, USA*, *89*(20), 9666–9670.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192.

de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, *70*(5), 789–794.

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology*, *62*(9), 1716–1722.

Esins, J., Schultz, J., Stemper, C., Kennerknecht, I., & Bülthoff, I. (2016). Face perception and test reliabilities in congenital prosopagnosia in seven tests. *i-Perception*, *7*(1), 2041669515625797.

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.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753.

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.

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics*, *72*(7), 1825–1838.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*(1), 98–106.

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics*, *75*(2), 278–286.

Ji, L., Chen, W., & Fu, X. (2014). Different roles of foveal and extrafoveal vision in ensemble representation for facial expressions. In D. Harris (Ed.), *International conference on engineering psychology and cognitive ergonomics* (pp. 164–173). Cham, Switzerland: Springer International Publishing AG.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36* ECVP Abstract Supplement.

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]

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *The Journal of Neuroscience*, *7*(11), 3416–3468.

Morgan, M. J., & Glennerster, A. (1991). Efficiency of locating centres of dot-clusters by human observers. *Vision Research*, *31*(12), 2075–2083.

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788.

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.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Pestilli, F., & Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Research*, *45*(14), 1867–1875.

Schiller, P. H., Logothetis, N. K., & Charles, E. R. (1990). Role of the color-opponent and broad-band channels in vision. *Visual Neuroscience*, *5*(4), 321–346.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*(5), 402–409.

Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, *94*(11), 1948–1962.

Thornton, I. M., Srismith, D., Oxner, M., & Hayward, W. G. (2014). Establishing a majority: Observer race influences estimates of crowd ethnicity. *i-Perception*, *5*(5), 490.

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research*, *32*(5), 931–941.

Watamaniuk, S. N., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research*, *29*(1), 47–59.

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *ACM SIGGRAPH Computer Graphics*, *18*(1), 24.

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]

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84.