Precise ensemble face representation given incomplete visual input

Jason M Haberman[1] & Lauren Ulrich[2]

[1]The Department of Psychology, Rhodes College, Tennessee, USA

[2]College of Health Sciences, William Carey University, Mississippi, USA

To whom correspondence should be addressed:

Jason Haberman

2000 N Parkway

Rhodes College

habermanj@rhodes.edu

Abstract

Humans can recognize faces in the presence of environmental noise. Here, we explore whether ensemble perception of faces is similarly robust. Is summary statistical information available from crowds of faces that are visually incomplete? Observers viewed sets of faces varying in identity or expression and adjusted a test face to match the perceived average. In one condition, faces amodally completed behind horizontal bars. In another condition, identical facial information was presented, but in the foreground (i.e., face parts appeared on fragmented strips in front of a background). Baseline performance was determined by performance on sets of fully visible faces. The results revealed that the ensemble representation of amodally completing sets was significantly better than the fragmented sets, and marginally worse than in the fully visible condition. These results suggest that some ensemble information is available given limited visual input, and supports a growing body of work suggesting that ensembles may be represented in the absence of complete visual information.

Keywords: Ensemble perception, faces, amodal completion

At a given moment, our retinas are inundated with millions of bits of information —

more information than the brain can consciously represent at any one time. A host of studies on

change blindness and capacity limitations of attention verify this, pointing to an extremely sparse

conscious visual experience (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997; Rensink, ORegan,

& Clark, 1997; Simons & Chabris, 1999; Treisman, 1982). Despite these limitations in conscious

visual awareness, the scene beyond the focus of attention does not altogether vanish; *something* is

available, and that something may aide in the deployment of limited attentional resources

(Wolfe, 1994; Wolfe & Horowitz, 2004). It has been proposed that, in the face of overwhelming

information, the visual system can exploit the redundancy of natural scenes by representing their

summary statistics (Alvarez, 2011; Haberman & Whitney, 2012; Whitney, Haberman, & Sweeny,

2014), a phenomenon known as ensemble perception (Ariely, 2001). Ensemble perception is a

robust and ubiquitous heuristic, operating efficiently (Alvarez & Oliva, 2009; Chong & Treisman,

2003; Haberman & Whitney, 2011) across all levels of the visual system (Haberman, Brady, &

Alvarez, 2015). It is independent of single object recognition (Chong, Joo, Emmanouil, &

Treisman, 2008; Whitney et al., 2014), functional under diminished attentional resources

(Alvarez & Oliva, 2008), and sensitive to an array of summary statistics beyond the central

moment (Dakin, 1999; Haberman, Lee, & Whitney, 2015; Solomon, 2010). Ensembles may be

derived across a variety of visual domains, from oriented gabors (Attarha & Moore, 2015; Dakin,

Bex, Cass, & Watt, 2009; Dakin & Watt, 1997; Parkes, Lund, Angelucci, Solomon, & Morgan,

2001) to faces varying in expression and identity (Fockert & Wolfenstein, 2009; Haberman &

Whitney, 2007, 2009; Leib et al., 2014), and many of these representations appear to be

mechanistically independent (Haberman et al., 2015).

Demonstrations of ensemble representations often reflect a warped memory trace, whereby observers recall summary information not ever displayed rather than individual object information actually presented (e.g., Ariely, 2001; Haberman & Whitney, 2009; Maule, Witzel, & Franklin, 2014). This may be regarded as the visual system's attempt to efficiently and accurately summarize a complex visual scene. For example, when asked whether a test item appeared in the previously displayed set, observers will false alarm to the average item (e.g., average expression) while showing relative little awareness of the individual items composing that set (Ariely, 2001; Haberman & Whitney, 2009). Thus, it appears the visual system automatically derives information that is not explicitly displayed, a testament to the power of the ensemble heuristic. This flexibility raises an important question: how much physical information is needed to generate an accurate summary representation?

This question gets at the notion of an *amodal* ensemble — summary information in the absence of physical input. If observers can generate a precise summary representation in the presence of limited information (due to occlusion or other interference), it would support the oft-made claim that ensembles provide visual stability in a noisy and dynamic environment (Cohen, Dennett, & Kanwisher, 2016; Corbett & Melcher, 2014; Whitney et al., 2014). That is, despite impoverished visual information, ensemble perception might continue to provide veridical and valuable information about natural scenes.

There already exists some evidence that ensemble information remains available despite low-fidelity input, suggesting mechanisms are in place to overcome environmental interference (e.g., occlusion). For example, even when foveal information is obscured from view, observers can still derive precise ensemble information based on noisy peripheral input (Wolfe,

Kosovicheva, Leib, Wood, & Whitney, 2015). Additionally, observers can discriminate average

expression information when they are unable to localize individual changes driving the average

differences (Haberman & Whitney, 2011) and even when they are unable to discriminate

individual expressions due to crowding (Fischer & Whitney, 2011). All this points to the

possibility that ensembles may be represented amodally, that is, summary information may be

available even in the absence of physical information.

In the current set of experiments, we explored amodal ensemble representation by

leveraging what is already known about amodal completion, a strong cue to occlusion and scene

depth (e.g., Coren, 1972; Kanizsa, 1976). In amodal completion, an object appears to complete

behind an occluding surface — imagine viewing a face that is behind a set of blinds (e.g., Figure

1). Various Gestalt cues drive this percept (e.g., good continuation, good form; Wertheimer,

1923), countering the possible and valid percept of multiple discrete and discontinuous objects.

With amodal completion, object identification proceeds unimpaired, despite the reduction of

available information; the expression of the face depicted on the left of Figure 1 is no less

identifiable than the complete, identical face on the right. Thus, amodal completion offers an

elegant approach to exploring whether precise ensemble information, like individual object

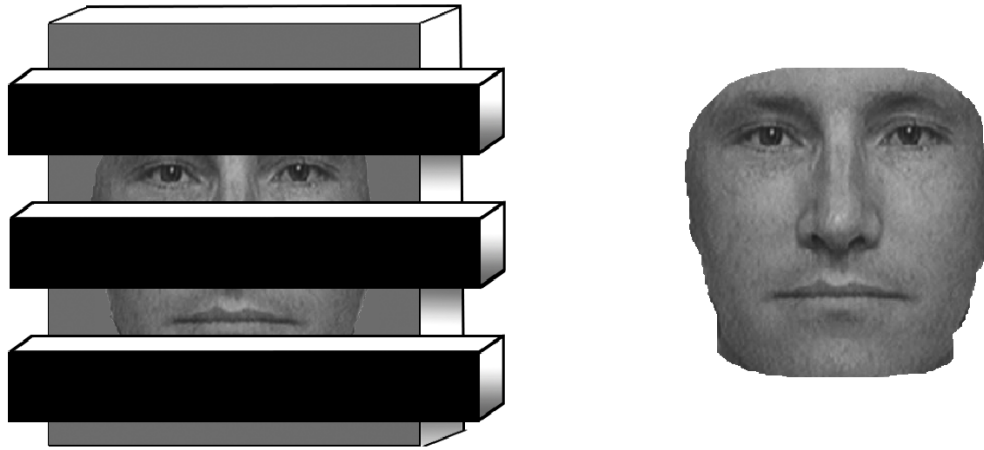perception, may be derived under impoverished viewing conditions.

Figure 1: Both faces are equally identifiable, even though the left image is partially occluded by a series of blinds.

Ensemble perception is a hierarchical operation, whereby different visual domains are represented by independently operating mechanisms (Haberman et al., 2015) — in this study we chose to focus on the domain of face perception. Ensemble face perception is surprising given the computational sophistication required for facial processing. While several studies have revealed robust support for ensemble face perception (e.g., Fischer & Whitney, 2011; Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007; Leib et al., 2014; Leib et al., 2012; Sweeny, Grabowecky, Paller, & Suzuki, 2009), it may nonetheless rely upon having access to complete and uninterrupted visual information. If the ability to represent the average expression or the average identity of a set of faces remains unimpaired given impoverished information, however, it would suggest a critical role for the ensemble heuristic in perceptual stability. While advanced artificial visual systems still struggle to properly identify faces in the presence of noise or occlusion (Scheirer, Anthony, Nakayama, & Cox, 2014), the human visual system might overcome these perceptual limitations by *combining* noisy information across a complex scene.

Experiment 1A

In Experiment 1A, we explored the representation of average identity when crowds of faces were visually incomplete. A significant body of research already suggests we can derive high-level identity information (e.g., Fockert & Wolfenstein, 2009; Haberman et al., 2015; Neumann, Schweinberger, & Burton, 2013), but how much does this depend on having holistic information available to the visual system? We presented observers with sets of faces varying in identity and asked them to report the average identity of a given set. In some of the conditions, faces were partially occluded or partially fragmented (Figure 2). It is already known that face processing (e.g., recognition) is robust to environmental occlusion or noise (e.g., Nakayama, He, & Shimojo, 1995; Sinha, Balas, Ostrovsky, & Russell, 2006), even when key information is obscured from view. Here we tested whether ensemble processing of partially occluded faces is also robust to visual obstruction. If the ensemble representation of occluded faces does not suffer relative to normal faces, it would suggest ensembles can be represented given incomplete visual information. In addition, it would add to the body of work suggesting that face averaging is a high-level, holistic process, and does not depend on piecemeal, featural analysis or high-fidelity individual item representation (e.g., Haberman & Whitney, 2010; Leib et al., 2014; Rhodes et al., 2018).

Included in this experiment is a critical condition that forces a part-based analysis of sets of faces. Instead of occluders blocking parts of each face, fragments of each face were displayed in the fore — the very same fragments that were part of the background in the occluded condition (Figure 2). In this condition, ensemble face processing should be disrupted, since the benefits of amodal completion are no longer available (Sekuler & Murray, 2001). Taken together, these

results would suggest that high-level ensemble processing may operate over inferred visual

information.

Method

Participants

Thirty-eight Rhodes College undergraduates, ages 18 to 22, participated in this study for

either course credit or monetary compensation. The compensation rate was $10 per hour. All

participants gave informed consent and had normal or corrected-to-normal vision. This

research, and all research described herein, was approved by and conducted in accordance with

the Institutional Review Board at Rhodes College.

Stimuli and Design

Observers were presented with sets of faces varying in identity. Stimuli consisted of 360

linearly interpolated identity morphs, taken from the Harvard Face Database, of three distinct

male faces (A-B-C-A), generated using MorphAge software (version 4.1.3, Creaceed). Face

morphs were nominally separated from one another in identity units, with each unit

corresponding to a degree in morph space. Face identity formed a circular stimulus space

spanning 360°. All stimuli in this and future experiments were presented in grayscale using

custom scripts developed in Psychophysics toolbox (Brainard, 1997) within Matlab (Mathworks,

Natick, MA).

Three conditions were included: Sets of faces behind occluding bars (behind), sets of face

fragments presented in the foreground (front), and sets of fully intact faces (standard; see Figure

2). The face pieces presented in the behind and front conditions contained identical face

information, only the figure ground relation was flipped. Each face in the standard condition was

displayed at 180×180 pixels, subtending 5.2°×5.2° of visual angle. Each face in the behind and

front conditions were 259×259 pixels, subtending 7.5° x 7.5° of visual angle (note that while the

overall size of the image in the standard and front/behind conditions differed, this was done to

equate the relative size of the visible facial information on the screen). Sets were followed by a

single face in the center of the screen (described below) at 200×200, subtending 5.8°×5.8° of

visual angle.

A subset of these observers (n = 21) also participated in a fourth condition in which they

viewed sets of faces in the behind condition, but inverted (i.e., bars appearing in front of upside

down faces). Inverting faces is known to disrupt configural processing (Tanaka & Farah, 1993;

Young, Hellawell, & Hay, 1987), and because precise ensemble face perception relies upon such

configural information (Haberman & Whitney, 2007, 2009), an inverted ensemble condition

serves as a reasonable performance comparison.

Observers were presented with sets of four faces varying in identity (±13 and ±36 identity

units from the mean — this was based on pilot experiments that determined an average JND of

approximately 26 identity units for these face morphs). The mean of each set was randomly

selected on every trial, and then the set of items were centered on that mean.
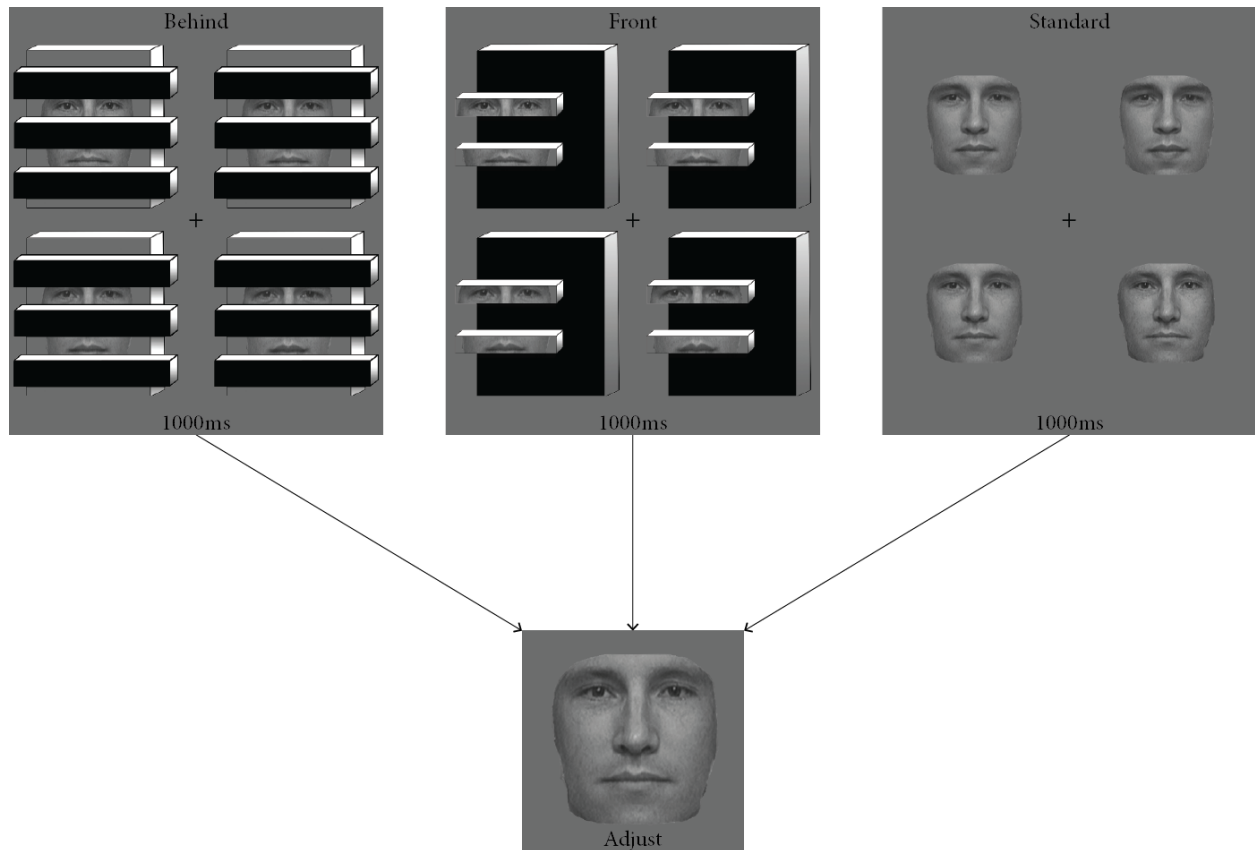
Figure 2: Conditions and procedure for Experiment 1. Observers viewed randomly interleaved conditions and adjusted a test face to match the mean identity of the preceding set. **Not shown is a fourth condition, given to 21 participants, composed of inverted faces in the behind condition.**

Procedure

On each trial, observers were tasked to report the average identity of a group of faces. Observers, with their heads resting on a chinrest 63 cm from the screen, viewed sets of four faces varying in identity for one second. After a 250 ms ISI, a single test face appeared at the center of the screen. The test face was always the original, unobstructed version, regardless of the condition of the preceding set. The starting identity of the test face was chosen at random from the identity wheel. Observers adjusted the test face to match the average identity of the preceding set by moving the mouse along the x-axis. As the mouse was moved, the appearance of the face

scrolled through the identity wheel. When observers were satisfied with their selection, they

pressed the spacebar to lock in their response and begin the next trial.

Prior to the beginning of the experiment, observers performed 24 practice trials in the

standard condition, which were discarded from further analysis. The primary task consisted of 80

trials in each of the three conditions, for a total of 240 trials. For those participants who also

viewed the inverted faces, there were 60 trials for each condition, for a total of 240 trials.

Data Availability

All data generated or analyzed for these experiments are available from the corresponding

author upon request.

<div align="center">Results and Discussion</div>

For each observer and condition, the mean absolute error was calculated as an index of

average identity precision (i.e., how far away observer responses were from the actual mean of

the set). Smaller absolute error indicates greater ensemble precision. Observer performance (as

indicated by mean absolute error) that was 2 standard deviations worse than the average overall

performance on any condition was excluded from analysis, resulting in the exclusion of 2

observers, for a total of 36 participants.

A one-way repeated measures ANOVA revealed a significant effect of condition ($F(2, 74)$

$= 18.1$, $p < .0001$; $\eta = .33$; Figure 3), where error for the standard condition was smallest ($M_{standard}$

$= 48.8°$), error for the behind condition was in the middle ($M_{behind} = 53.1°$), and error for the front

condition was highest ($M_{front} = 58.0°$). Tukey HSD tests revealed that all conditions significantly

differed from one another. Although both behind and front conditions differed from the

standard condition, the magnitude of the difference for the front condition, where observers did

not derive the benefit of amodal completion, was nearly twice that of the behind condition. Thus, while partial removal of facial information did negatively impact ensemble performance, this was exacerbated when the figure ground relation was reversed. This experiment reveals that ensemble perception is robust enough to overcome some, but not all, of the effects of the physical blockage of parts of the face. These deleterious effects are mitigated through configural processing, which is facilitated by amodal completion in the behind condition (Nakayama et al., 1995).

Interestingly, as much as performance suffered in the front condition, it was still significantly better than in the inverted condition (when directly comparing the observers who viewed both conditions; M_{front} = 59.9°; $M_{inverted}$ = 69.5°; $t(19)$ = -3.31, p = .004, η = .37). Thus, judgments in the front condition are not at floor. Switching the face fragments from background to foreground clearly disrupts configural processing, but not to the same extent that inverting them does.
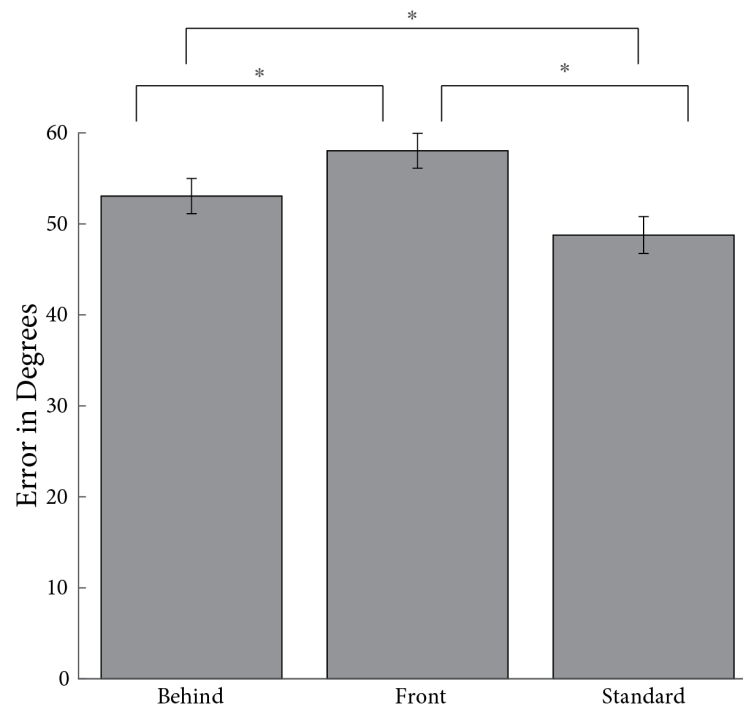

Figure 3: Average absolute error for the three ensemble identity conditions for Experiment 1. Observers were significantly worse in the front condition relative to the standard condition, and performance in the behind

condition similarly suffered relative to the standard condition. * indicates p < .05. Error bars indicate standard error of the mean (SEM).

Experiment 2

Experiment 1 established that ensemble expression information is available even if parts of the faces are blocked from view, although performance suffered to some extent. It is tempting to conclude that people are performing the ensemble calculus over the missing information in their average representation, albeit at reduced precision relative to the standard condition. However, the previous experiment did not explicitly establish *what* information individuals were averaging. In the current experiment, observers viewed the same conditions as described in Experiment 1, but they adjusted a test face that only contained the complementary facial information (Figure 4). Only parts that were missing from the original stimulus were shown during the adjustment phase, thereby testing the quality of the representation of the missing facial information.

Method

Participants

Fifteen Rhodes College undergraduates, ages 18 to 22, participated in this study for either course credit or monetary compensation. The compensation rate was $10 per hour. All participants gave informed consent and had normal or corrected-to-normal vision.

Stimuli, Design, and Procedure

Setup for this experiment was nearly identical to that described for Experiment 1. The one major difference was in the test face displayed during the adjustment portion of the trial. Instead of adjusting an intact face to match the perceived mean of the preceding set, observers adjusted a test face containing the complementary missing information (to maintain consistency,

this was also true for the standard sets where all information was visible). In our stimuli,

forehead, nose, and chin information was absent in the behind and front conditions, and thus

that was what was visible to observers during the adjustment phase (Figure 4).
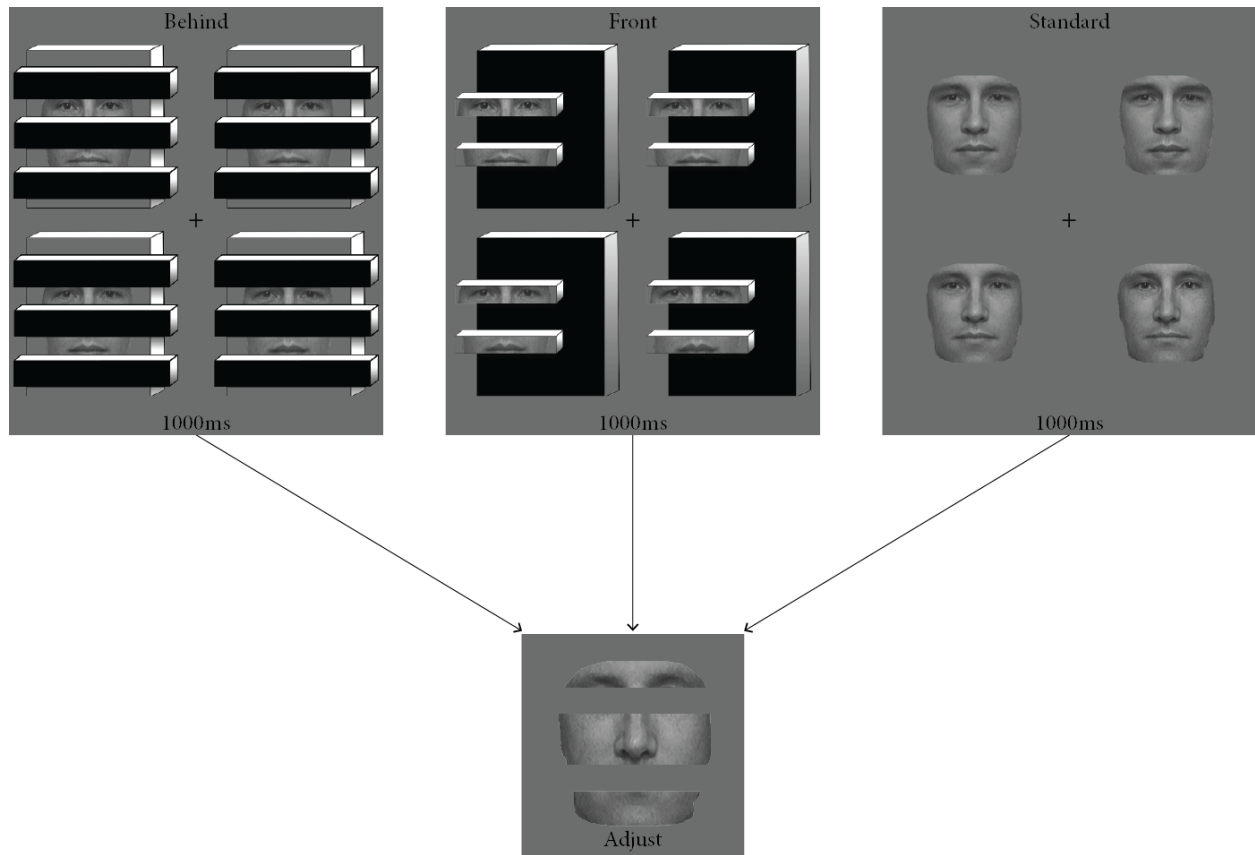

Figure 4: Conditions and procedure for Experiment 2. Observers viewed randomly interleaved conditions and adjusted a fragmented test face containing complementary missing parts to match the mean identity of the preceding set.

Results and Discussion

Results were analyzed as described in Experiment 1. Observers whose performance was

two standard deviations worse than the average performance for any condition were excluded

from analysis, resulting in the exclusion of three observers, for a total of 12 participants.

Results of this experiment are displayed in Figure 5. There was a significant effect of

viewing condition, as revealed by a one-way repeated-measures ANOVA ($F(2, 22) = 3.91$, $p = .04$,

η = .26). Performance appears to have suffered to the same extent in both the behind and front

conditions relative to the standard conditions (although a post-hoc Tukey HSD test revealed no

differences among conditions: M_{behind} = 64.8; M_{front} = 65.1; $M_{standard}$ = 58.5;). These results suggest

that when information is obscured from view, individuals are not explicitly averaging the missing

information, or are doing so poorly. Amodal completion may not actually allow the visual system

to recreate the missing information, but rather support a 'best-guess' heuristic, akin to visual

completion. Performance in the standard condition may be marginally better than the front and

behind conditions because observers do not have to visually complete any information. Although

note that performance even in the standard condition suffered in this task relative to Experiment

1, as revealed by a between-subjects t-test ($t(48)$ = -2.25, p = .03), because the relative amount of

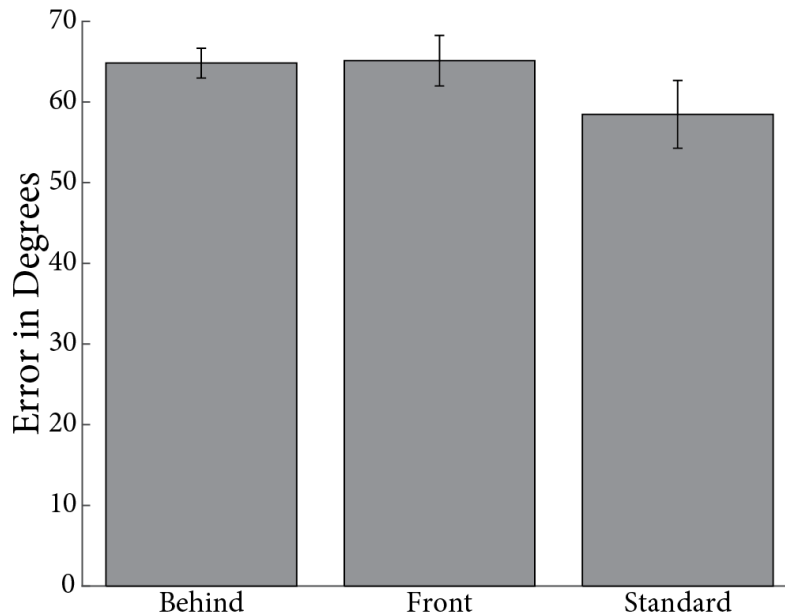information available in the test stimulus itself is reduced.

Figure 5: Average absolute error for the three ensemble identity conditions when observers adjusted complementary missing face information **for Experiment 2**. There was a significant effect of condition at the p < .05 level. Error bars indicate standard error of the mean.

Experiment 3

This experiment generalizes the findings established in Experiment 1 by exploring high-level, amodal ensemble representations for sets of faces varying in expression. Although substantial evidence points to robust ensemble perceptual abilities for both identity and emotion (Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2010), one cannot assume equivalence given their well-established behavioral and neural independence (Bruce & Young, 1986; Haxby & Gobbini, 2011).

Method

Participants

Twenty Rhodes College undergraduates, ages 18 to 21, participated in this study for either course credit or monetary compensation. The compensation rate was $10 per hour. All participants gave informed consent and had normal or corrected-to-normal vision.

Stimuli and Design

Observers were presented with sets of faces that differed in emotional expression (Figure 6). In this experiment, the face morphs constituted a circle of 360 expressions (as with the identity stimuli), spanning from angry to happy to sad and back to angry. Faces came from the publicly available Karolinska Directed Emotional Faces database (KDEF)(Lundqvist, Flykt, & Öhman, 1998). Distance between two given faces was nominally measured in emotional units. As in Experiment 1, there were three conditions: Behind, front, and standard. Faces in the behind and front conditions were 292×292 pixels in size, subtending 8.5°×8.5° degrees of visual angle. Faces in the standard condition were 122×158 pixels in size, subtending 3.5°×4.6° of visual angle. The subsequent test face was 121×158, subtending 4.4°×5.7° of visual angle.

As described previously, observers were presented with sets of four faces varying in emotional expression (±13 and ±39 emotional units from the mean). The mean of each set was randomly selected on every trial, and then the set items were centered on that mean.

Procedure

The task was nearly identical to that described in Experiment 1A, but instead of reporting the average identity, observers reported the average expression (Figure 6). As in the previous experiment, observers viewed each set for one second, followed by a single test face after a 250 ms ISI. Observers adjusted the test face to match the average expression of the preceding set of faces.

Prior to beginning the experiment, observers performed 24 practice trials in the standard condition, consistent with the number from the previous experiments. For the primary task, observers performed 80 test trials in each of the three conditions, for a total of 240 test trials.
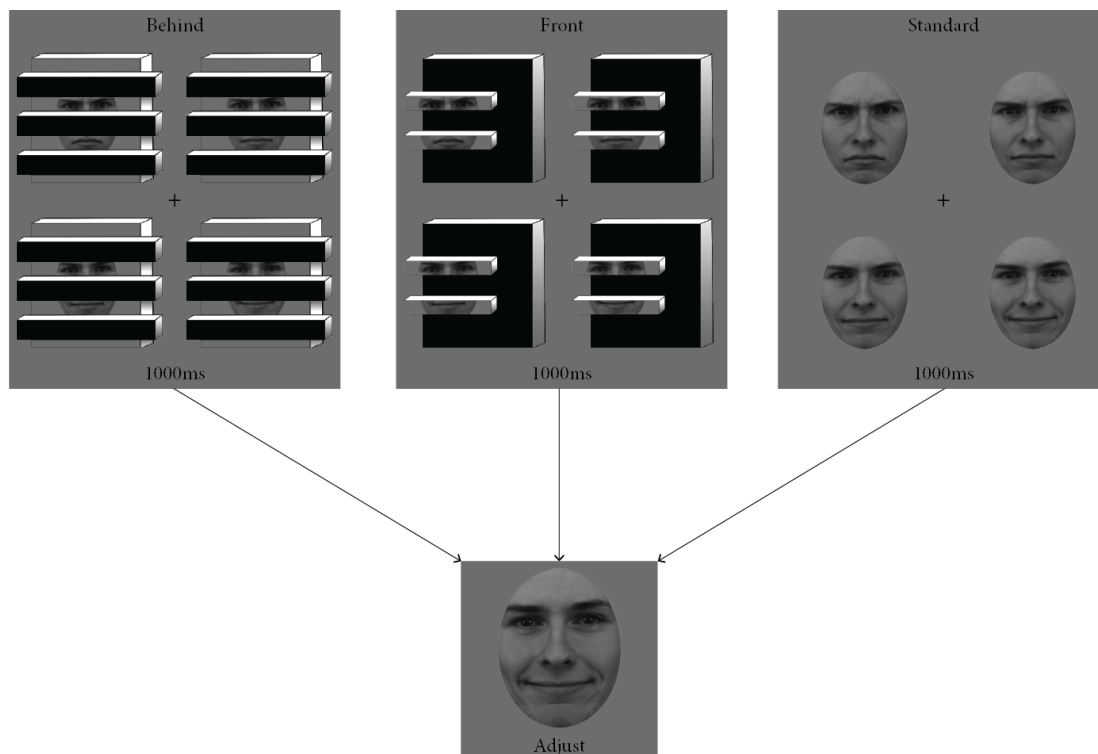

Figure 6: Conditions and procedure for Experiment 3. Observers viewed randomly interleaved conditions and adjusted a test face to match the mean expression of the preceding set.

Results

Performance was assessed as previously described. Observers whose performance was two standard deviations worse than the average performance for any condition were excluded from analysis, resulting in the exclusion of three observers, for a total of 17 participants.

Consistent with the previous experiment, a one-way repeated measures ANOVA revealed a significant effect of condition ($F(2, 32) = 13.7$, $p < .0001$; $\eta = .11$; Figure 7), where error for the standard condition was smallest ($M_{standard} = 42.2°$), error for the behind condition was in the middle ($M_{behind} = 44.7°$), and error for the front condition was highest ($M_{front} = 49.7°$). A Tukey HSD test revealed the front condition was significantly worse than both the standard and the behind conditions ($p < .01$), consistent with the pattern described in Experiment 1. The post-hoc test showed no difference between the standard and behind conditions. However, this does not mean these two conditions are equivalent, only that the benefits of amodal completion extend to processing crowds of emotionally varying faces. It also supports the notion that high-level ensembles may be represented even given an incomplete visual stimulus. Ensemble representation ability is better when the faces amodally complete than when they do not, perhaps facilitated by easier visual completion.
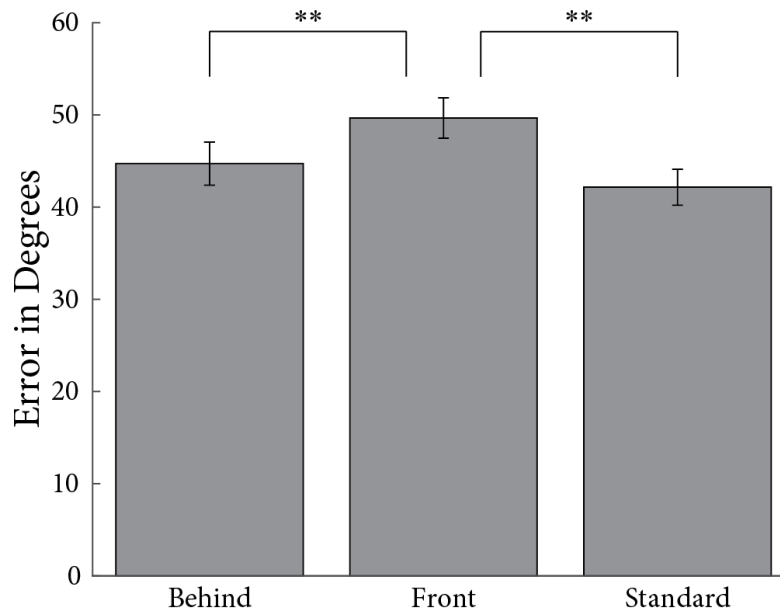
Figure 7: Average absolute error for the three ensemble expression conditions **for Experiment 3**. Observers were significantly worse in the front condition relative to the standard and behind conditions, while there was no difference in performance between the behind and standard conditions. ** indicates p < .01. Error bars indicate standard error of the mean.

General Discussion

These experiments reveal that the visual system can extract high-level ensembles even in the presence of incomplete scene information, but at some cost. Observers reported the average expression or average identity for sets of faces that amodally completed behind several occluding bars. The precision of amodally completing sets was only marginally worse for both stimulus domains than when the entirety of the stimulus was presented. This was not a floor effect, as switching the figure ground relationship such that the face fragments were in front of the bars (e.g., Figures 2 and 7) further disrupted ensemble performance. Although placing the face fragments in the fore negatively impacted ensemble performance, inverting the amodally completing face stimuli made performance even worse, suggesting that noisy ensemble information was still available from the face fragments.

Finally, Experiment 2 suggested that missing ensemble information obscured by the bars may be partially derived, albeit less precisely. Given the reduction in precision, it reveals that the representation is not based explicitly on the information obscured from view — when observers had to adjust a test face with the complementary missing information to match the mean of the preceding set, performance suffered in both the behind and front conditions. Thus, it does not appear that observers are explicitly representing the missing information, but are rather making their inferences as to what was obscured. Given that even basic stimuli, such as familiar shapes or objects, are often variably completed (Boselie, 1988; Van Lier, 1999), it follows that amodally completing faces, where there are even more degrees of freedom of interpretation, are somewhat less precise than fully visible faces (although note, performance was not at floor, as performance on inverted stimuli was still substantially worse).

In the standard condition, observers may have been able to rely upon memory traces of the intact faces to better adjust the fragmented test face. However, adjusting the fragmented test face also negatively impacted the performance in the standard condition relative to when the entire test face was visible. Impairment may also be a function of the relatively less important facial features observers had to visually complete. Most observers, when viewing a face, rely upon specific critical features for identification or emotional recognition (Schyns, Bonnar, & Gosselin, 2002), features that may have been obscured in our task. Thus, deriving a high-fidelity representation of secondary features such as the brow or chin may be of less importance. Future studies should manipulate the kinds of features that amodally complete in the set to test whether amodally completing information critical to identification or emotion recognition may be explicitly represented.

The visual system is often presented with noisy versions of objects in natural scenes, thus in order for object recognition to function it must be robust to common sources of interference. It may not be surprising, then, that one's ability to recognize a singular face or other object is unimpaired even when it is partially occluded (Nakayama et al., 1995; Sinha et al., 2006). Our results add to this finding, showing only marginal cost to judging sets of partially occluded faces. Overall, this finding suggests that the visual system can infer the ensemble based on exemplar information that is only partially visible.

As noted, viewing fragments in the fore disrupted the ensemble calculus. While observers still had limited access to the ensemble information, the performance decrement suggests a distinct cognitive process from the one operating when faces were amodally completing behind the bars. In other words, the 2D image fragments, while necessary, are insufficient to generate a high-level ensemble representation on their own. This view is consistent with the finding that face and body ensembles are invariant to viewpoint differences (Leib et al., 2014; Sweeney, Haroz, & Whitney, 2012), which would not be possible if only fragmented features were visible. The overall performance benefit in the amodal condition makes sense in light of how often we encounter faces: Behind objects and partially occluded, a situation much more likely to occur than fragmented face pieces.

These experiments dovetail with recent dissertation work examining occluded ensembles for low-level visual stimuli (Lee, 2015). They represent an early step in understanding what facial components are necessary to generate an accurate ensemble. Ensemble performance suffered a bit in the amodal condition perhaps because too much information was obscured from view — it would be worthwhile to systematically occlude various features of the set to see which elements

most affect ensemble representations. Predictions could emerge based on the body of work

showing what information is most critical for transmitting a given facial expression (Smith,

Cottrell, Gosselin, & Schyns, 2005).

These findings contribute to a growing body of work suggesting ensembles provide a

source of stability given the limitations of visual consciousness (Alvarez, 2011; Alvarez & Oliva,

2009; Cohen et al., 2016; Fischer & Whitney, 2011; Haberman & Whitney, 2011). For example,

ensemble perception operates rapidly, in as little as 50 ms (Haberman & Whitney, 2009), is

generally unaffected by set size (Chong & Treisman, 2003), and occurs beyond the focus of

attention (Alvarez & Oliva, 2008), all useful features for a system trying to compensate for limited

conscious access. Add to this list the current set of experiments, which suggest some ensemble

information, albeit noisy, is derived even when visual information is altogether missing —

arguably a critical component in the effort to create the impression of a complete and holistic

visual experience.

Conclusion

This is one of the first demonstrations that ensemble information may be generated

amodally. Furthermore, these amodal representations were revealed in a distinctly high-level

example. These results lead one to question the extent to which amodal ensemble representations

operate. Future experiments should test other domains within the visual hierarchy, as the system

often encounters objects in less-than-ideal viewing conditions. Given the ubiquitous scope and

robust nature of ensemble perception, it might be uniquely situated to overcome the challenges

of such impoverished (and typical) visual scenes.

Additional Information

Author Contributions

J. Haberman developed the study concept. All authors contributed to the study design. Testing and data collection were performed by L. Ulrich. J. Haberman and L. Ulrich performed the data analysis and interpretation. L. Ulrich drafted early versions of the methods and results sections. J. Haberman revised the methods and results sections and wrote the introduction, discussion, and conclusion. All authors approved the final version of the manuscript for submission.

Competing Financial Interests

The authors declare no competing financial interests.

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., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106-111. doi:10.1111/j.0963-7214.2004.01502006.x

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

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proc Natl Acad Sci U S A, 106*(18), 7345-7350.

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

Attarha, M., & Moore, C. M. (2015). The capacity limitations of orientation summary statistics. *Attention, Perception, & Psychophysics, 77*(4), 1116-1131.

Boselie, F. (1988). Local versus global minima in visual pattern completion. *Perception & Psychophysics, 43*(5), 431-445.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Bruce, V., & Young, A. (1986). Understanding face recognition. *Br J Psychol, 77 (Pt 3)*, 305-327.

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; discussion 1335-1326. doi:10.3758/PP.70.7.1327

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision research, 43*(4), 393-404.

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in cognitive sciences, 20*(5), 324-335.

Corbett, J. E., & Melcher, D. (2014). Characterizing ensemble statistics: Mean size is represented across multiple frames of reference. *Attention, Perception, & Psychophysics, 76*(3), 746-758.

Coren, S. (1972). Subjective contours and apparent depth. *Psychological review, 79*(4), 359.

Dakin, S. C. (1999). Orientation variance as a quantifier of structure in texture. *Spatial Vision, 12*(1), 1-30.

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). doi:Artn 28

Doi 10.1167/9.11.28

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

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of neurophysiology, 106*(3), 1389-1398.

Fockert, d., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Q J Exp Psychol (Colchester)*, 1-7.

Haberman, 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.

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a

    crowd of faces. *Journal of Vision, 15*(4), 16-16.

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.

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change

    localization fails. *Psychonomic Bulletin & Review, 18*(5), 855-859.

Haberman, J., & Whitney, D. (2012). From Perception to Concsciousness: Searching with Anne

    Treisman. In J. Wolfe & L. Robertson (Eds.), (pp. 339-349): Oxford University Press.

Haxby, J. V., & Gobbini, M. I. (2011). Distributed neural systems for face perception. In A. J.

    Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face

    perception* (pp. 93-107).

Kanizsa, G. (1976). Subjective contours. *Scientific American, 234*(4), 48-52.

Lee, J.-E. J. (2015). Statistical processing: mean size perception of partially occluded sets.

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.

Leib, A. Y., Puri, A. M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd

    perception in prosopagnosia. *Neuropsychologia, 50*(7), 1698-1707.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and

conjunctions. *Nature, 390*(6657), 279-281.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF).

CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska

Institutet, 91-630.

Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: metric and

categorical effects on ensemble perception of hue. *JOSA A, 31*(4), A93-A102.

Nakayama, K., He, Z., & Shimojo, S. (1995). Invitation to cognitive science.

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and

individual identity from sets of famous faces. *Cognition, 128*(1), 56-63.

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.

Rensink, R. A., ORegan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to

perceive changes in scenes. *Psychological Science, 8*(5), 368-373.

Rhodes, G., Neumann, M., Ewing, L., Bank, S., Read, A., Engfors, L. M., . . . Palermo, R. (2018).

Ensemble coding of faces occurs in children and develops dissociably from coding of

individual faces. *Developmental science, 21*(2), e12540.

Scheirer, W. J., Anthony, S. E., Nakayama, K., & Cox, D. D. (2014). Perceptual annotation:

Measuring human vision to improve computer vision. *IEEE transactions on pattern*

analysis and machine intelligence, 36(8), 1679-1686.

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.

Sekuler, A. B., & Murray, R. F. (2001). Amodal Completion: A Case Study in Grouping. In T. F. Shipley & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (Vol. 130, pp. 265-293). The Netherlands: Elsevier.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059-1074.

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.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16*(3), 184-189.

Solomon. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *J Vis, 10*(14), 19. doi:10.14.19 [pii]

10.1167/10.14.19

Sweeney, T. D., Haroz, S., & Whitney, D. (2012). Perceiving Group Behavior: Sensitive Ensemble Coding Mechanisms for Biological Motion of Human Crowds. *Journal of Experimental Psychology: Human Perception and Performance*. doi:10.1037/a0028712

Sweeny, T. D., Grabowecky, M., Paller, K., & Suzuki, S. (2009). Within-hemifield perceptual averaging of facial expression s predicted by neural averaging. *Journal of Vision, 9*(3), 1-11.

Tanaka, J. W., & Farah, M. J. (1993). Parts and Wholes in Face Recognition. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 46*(2), 225-245.

Treisman, A. (1982). Perceptual Grouping and Attention in Visual-Search for Features and for

   Objects. *Journal of Experimental Psychology-Human Perception and Performance, 8*(2),

   194-214.

Van Lier, R. (1999). Investigating global effects in visual occlusion: From a partly occluded

   square to the back of a tree-trunk. *Acta Psychologica, 102*(2-3), 203-220.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische forschung, 4*,

   301-350.

Whitney, D., Haberman, J., & Sweeny, T. (2014). From textures to crowds: multiple levels of

   summary statistical perception. *The New Visual Neurosciences., JS Werner and LM

   Chalupa, eds.(MIT, 2014)*, 685-709.

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), 11-11.

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychon Bull Rev, 1*(2),

   202-238. doi:10.3758/BF03200774

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention

   and how do they do it? *Nat Rev Neurosci, 5*(6), 495-501. doi:10.1038/nrn1411

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational Information in Face

   Perception. *Perception, 16*(6), 747-759.