

**Perceiving the mean?  
Investigating spatial frequency as a constituent of high-level  
ensemble coding***Teun Beumer  
Master Thesis  
19-9-2021*

*Abstract*This paper investigated the relationship between ensemble coding, the ability of humans to readily report average parameters such as average colour, speed, orientation or emotion and spatial frequency defined by luminance variations cycling over different amounts of space. Some research regarding ensemble coding differentiates between low- and high-level ensemble coding, low-level ensemble coding being based on low-level features such as size, colour or spatial frequency and high-level ensemble coding being more holistic and research concerning this is primarily based on faces. However, as recent advancements in science have discovered that spatial frequency influences face perception it might be that high-level ensembles are actually ensembles based on the low-level feature spatial frequency. The analysis showed an ensemble coding effect but yielded inconclusive results regarding whether spatial frequency influenced this ensemble coding. This leads to the conclusion that high-level ensemble coding probably has a basis in holistic stimuli.

Applied Cognitive Psychology  
Utrecht University  
Supervisor: dr. Sjoerd Stuit  
Second supervisor: dr. Jeroen Benjamins  
Word count: 2653  
ECTS: 27,5  
Student number: 5954398  
*For further inquiries into this thesis don’t hesitate to contact me on teun44b@gmail.com*

*Introduction*

While walking through a forest, field or on a beach. What becomes evident is that the world is full of the same things: grains of sand, leaves, branches, blades of grass. The brain copes with this redundancy by perceiving the forest rather than perceiving every tree individually (Haberman & Whitney, 2007). This adaptive mechanism which derives statistical summary information from sets of similar objects or faces is called ensemble coding (Whitney, Yamanashi Leib, 2018). Ensemble coding makes it possible for humans to readily (within 100 milliseconds) report the motion, direction, speed, colour, hue, facial expression, orientation, family resemblance, gaze direction or size of objects in an ensemble (Whitney & Yamanashi Leib, 2018). Moreover, it is so efficient that humans can more accurately report the average value of a parameter (e.g. colour) than judge whether an object was present (Potter, 1976; Rensink, Regan & Clark, 1997; Ariely, 2001). Furthermore, Haberman, Lee & Whitney (2015) demonstrated that ensemble coding is more than taking the average, the researchers showed that participants were equally able to report the variance in emotions of a sample. Perceiving the variance gives insight into the estimated average, a group consisting only of angry faces looking your way means something different than a group displaying varying emotions. Additionally, some researchers discriminate between different levels of ensemble coding. Low-level ensemble coding has a computational basis in low-level cues, cues such as orientation, spatial position speed, hue, or size (Whitney & Yamanashi Leib, 2018; Watamaniuk & McKee 1998,Watamaniuk, Sekuler & Williams, 1989). Low-level ensemble coding of spatial frequency, colour and orientation might be responsible for texture recognition/discrimination and scene gist impressions (Kauffmann, Ramanoël & Peyrin, 2014; Landy, 2014; Oliva & Torralba, 2006). High-level ensemble coding, on the other hand, seems to have a basis in more holistic features, features which play a role in emotion recognition, gaze direction and gender- and identity identification (Haberman & Whitney 2009, 2007).   
 Interestingly, and in spite of ample research concerning ensemble statistics, it is not quite clear which visual stimuli the brain uses to create these ensembles. One theory suggests that high-level ensemble statistics are computed using both high- and low-level features, with these features both contributing to a holistic image (Haberman & Whitney, 2010; Whitney, Haberman & Sweeny, 2014; Han, Yamanashi Leib, Chen & Whitney, 2020). To address the question whether ensemble coding also occurs for higher-level visual cues such as faces without part-based cues such as identifiable features or surface texture characteristics, Han et al. (2020) conducted experiments with Mooney faces. Mooney faces are black and white shadow defined images that cannot be recognized in a part-based manner (Han et al, 2020). This research confirmed the integration of multiple Mooney faces into ensemble representations, concluding ensemble coding functions when holistic information is maximized and that higher-level ensemble coding is more than averaging individual features.  
 On the other hand, virtually all research into high-level ensemble coding is based on faces and recent research in the field of face perception yielded some interesting results. For decades there remained controversiality regarding which emotion elicits the most attentiveness. A study from Hansen and Hansen (1988) reported that angry faces are the most quickly noticed. A finding later reproduced by many studies (Lundqvist & Ohman, 2005; Lobue, 2009, Ceccarini & Caudek, 2013). All the while there were studies reporting a happiness superiority effect (Juth, Lundqvist, Karlsson & Öhman, 2005; Calvo & Nummenmaa, 2008; Hodsoll, Viding & Lavie, 2008). Eventually, in a later study by Savage and Lipp (2015) both these superiority effects were reproduced. But, more importantly, the researchers argued that the finding of different superiority effects might be because of differences in low-level visual cues. In fact, in later research one low-level visual cue in particular was found to be important: spatial frequency, defined by luminance variations cycling over different amounts of space (Stuit et al., 2021; Jeantet, Caharel, Schwan, Lighezzolo-Alnot & Laprevote, 2018). By isolating the spatial-structure information and spatial frequency content of emotional faces Stuit et al (2021) were able to provide evidence for the case that the perception of faces is heavily influenced by low-level features. This suggests that faces might not be the holistic high-level stimuli they appeared to be but rather are perceived through their low-level features.  
 To return to the field of ensemble statistics, it may be that just as in the case of face perception, what at first glance seems like high-level ensembles are actually ensembles consisting of low-level stimuli. To be more frank, what seems like the averaging of emotions on multiple faces might actually be the averaging of their spatial frequency content that produces ensemble perception. To test this hypothesis this paper proposes to examine spatial frequency as constituent for the ensemble perception of emotional faces. By biasing the perception of an ensemble using spatial frequency the aim is to influence the ensemble perception of participants. As higher spatial frequencies are linked to the identification of sad expressions and low spatial frequencies are linked to the identification of happy expressions this research will use the spatial frequency of those expressions to create a bias (Kumar & Srinivasan, 2011). The displays used for creating this bias contain happy or angry faces and noise patches containing the spatial frequency information of happy, angry or 1/f noise. If the noise patches containing the spatial frequency of happy faces influence ensemble coding, then the amount of angry faces necessary for a participant to designate the display as angry will be higher relative to the other conditions. Conversely, if the spatial frequency noise patches containing the spatial frequency information of angry faces influence ensemble coding then the amount of angry faces necessary for a participant to designate the display as angry will be the lowest relative to the other conditions. Finally, as a control, it is expected that the amount of angry faces necessary for designating the display as angry will be somewhere in between the other conditions when the display contain the 1/f spatial frequency noise patches.

*Methods*

Participants

49 people participated in the current study. Prior to the experiment participants filled out a consent form informing them their data would be anonymised. The study was approved by the ethical committee of the Utrecht University. Only the data of participants who completed the entire experiment was analysed.

Apparatus

Stimuli were created with an Apple Macbook Pro computer running OS X and Matlab 2019b with the psychophysics Toolbox extensions version 3.0. The experiment was coded in Inquisit 6. The experiment ran in InquisitPlayer, each participant was instructed on how to install it on their PC, the experiment could only be performed on a PC. After the experiment an explanation of how to uninstall the InquisitPlayer was shown.

Stimuli

The stimuli consisted of 480 displays. These displays contained 15 faces which could range from 0 angry or happy faces to 15 angry or happy faces, see Figure 1 for an example. Next to the images of faces the displays contained 30 noise patches. The noise patches were created by combining the Fourier magnitude spectrum of either happy faces or angry faces with a random phase spectrum. Each noise patch was based on a unique face. The noise patches consisted of happy noise, angry noise or 1/f noise patches. These Caucasian faces were sampled from the Radboud Faces database (Langner et al., 2010). Only faces with a frontal gaze were used and all faces were converted to greyscale.

*Afbeelding met stof

Automatisch gegenereerde beschrijving* **Figure 1.***An example of a display used in the experiment, the display contains 15 happy faces along with 30 noise patches containing the spatial frequency of either happy, angry or 1/f noise. Each display was shown for 100 milliseconds after which the participants were asked whether the perceived display was more angry or more happy.*

Design & Procedure

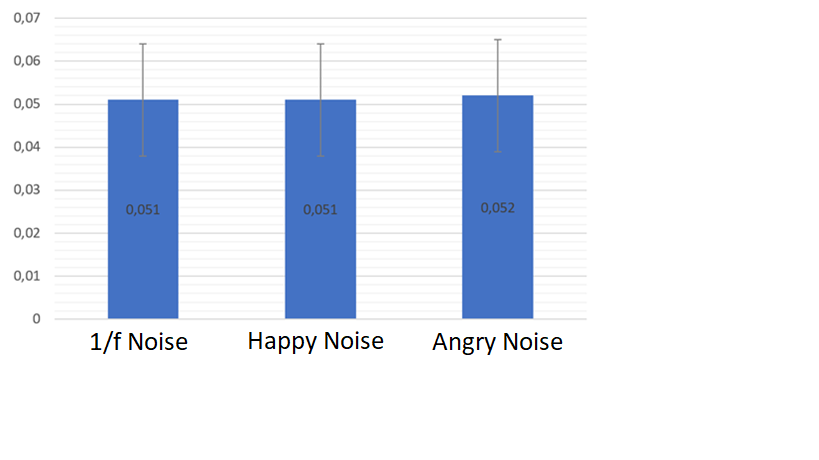
The experiment was performed online. Participants received written instructions on how to perform the experiment. Prior to the trials the participants were asked their age and sex. After which participants did 10 practice trials whereafter the experiment began. Participants were shown a white cross for 1000ms in the middle of their display as a prompt, after that the display was shown for 100ms followed by a black display with an instruction reminder asking them whether they perceived the display as more angry or more happy. Participants performed a two alternative forced choice task after the display was shown, they were instructed to press the “a” button if they thought the displays was more angry and they were instructed to press the “h” button if they thought the displays was more happy. The instruction reminder stayed on the screen until a response was recorded. Participants could stop at anytime by pressing “*ctrl +* q”.

Analyses

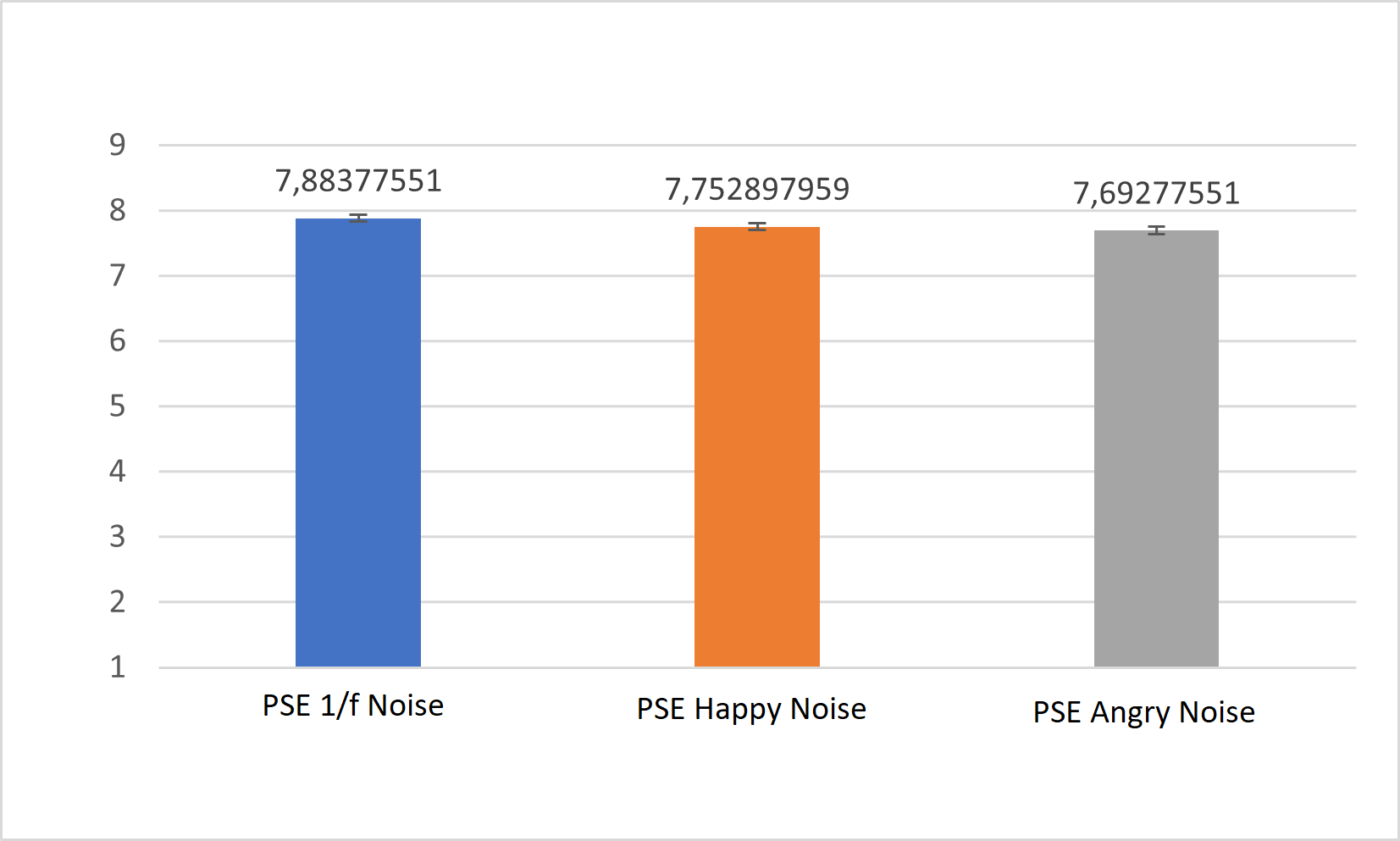
To establish whether there was any difference between the 1/f-noise, happy-noise and angry-noise conditions Points of Subjective Equality (PSE’s) were calculated. The PSE is the required number of angry faces in a display needed for a participant to give an angry response 50% of the time as estimated by a fitted regression model. This linear regression model was fitted over the data of each participant and each condition in order to calculate the regression slopes which were used in turn to compare the differences between the participant and per condition. A linear relationship between the amount of angry faces in the display and the percentage of an angry response is indicative of an ensemble effect. Analyses were conducted in JASP and SPSS.

*Results*

To test whether an ensemble effect showed in the data, for example whether more angry faces in the display would lead to more angry responses from the participants a one sample t-test was used. This one-sample t tests established the slopes of the regressions to be significantly bigger than 0. The means and standard deviations of the slopes for the 1/f noise, happy noise and angry noise are respectively: M = 0.051, SD = 0.012; M = 0.051, SD = 0.013; M = 0.052, SD = 0.013. The t tests of the 1/f noise, happy noise and angry noise are respectively: t(48) = 29.760, p < .001, d = 4.251; t(48) = 27.985, p < .001, d = 3.998; t(48) = 28.916, p < .001, d = 4.131. See figure 2.

**  
Figure 2.***A bar graph showing the means and standard deviations of the slopes of the simple regression per condition.*

To test whether the spatial frequency noise patches had an effect on the ensemble coding of emotion a Bayesian repeated measures analyses of variance was performed. The differences between the points of subjective equality (PSE’s) of the different noise types were investigated and the Bayesian ANOVA returned a BF of 0.908, ergo, an inconclusive result. See figure 3 for a representation of the data.

**Figure 3.***A bargraph showing the average PSE and error bars per condition.*

*Discussion*

This paper investigated the relationship between the ensemble coding of emotions and spatial frequency. In order to investigate whether spatial frequency influenced the ensemble coding of emotions the aim of the experiment was to create a bias. Using noise patches containing the spatial frequency of either 1/f noise, happy- or angry faces it was tried to influence the perception of the average emotions of a display. However, while the data showed an ensemble effect there was no conclusive evidence regarding whether the spatial frequency noise patches affected the participants’ ensemble coding. To sum up the parameters of the experiment: the sample was adequately large, there is evidence of spatial frequency affecting face perception, the time participants were shown the display should be adequate and displays are frequently used to establish ensemble coding. These parameters are nothing out of the ordinary and still no conclusive evidence was observed regarding whether ensemble coding is influenced by spatial frequency (Stuit et al, 2021; Whitney & Yamanashi Leib, 2018; Kumar & Srinivasan, 2011). Consequently, if spatial frequency influences the ensemble coding of emotions then the data would have been indicative of this.   
 As to why no conclusive evidence was detected there are various explanations. First of all, it might be that the ensemble coding of emotions is primarily a high-level process involving holistic stimuli. Faces have been designated as the ideal high-level stimuli for investigating high-level ensemble coding (Han et al. 2020;Whitney & Yamanashi Leib, 2018; Haberman & Whitney, 2010). Consequently is has been proven time and time again that people are able to report mean emotions, gender, gaze direction, family resemblance, head rotation and facial identity. All of these effects are based on the observations of faces and are also reported if the stimuli set consists of Mooney faces, black and white images which cannot be recognised in a feature-based manner (Han et al, 2020). Even more so, these effects were not observed if the faces were scrambled or inverted, leaving the features of faces intact but disorganising the familiar lay-out of a face (Haberman & Whitney 2009; Yamanashi Leib et al. 2012b). To put it another way, high-level ensemble coding most likely involves the perception of holistic stimuli and it seems more than likely that it does not lean on spatial frequency to give us immediate information on the status of the people in our direct surroundings. Still, it is interesting to compare this research with the research concerning face perception.   
 Even though this paper has found no conclusive evidence of spatial frequency being of importance to the ensemble perception of faces, recent research on face perception found that spatial frequency was a major influence on the perception of faces (Stuit et al. 2021). This seems paradoxical, there seems to be a big difference between the perception of one face and the perception of multiple faces. What should be kept in mind is perhaps the differences in the methods that were used. While high-level ensemble coding is primarily researched using displays with multiple faces, the research of Stuit et al. (2021) was performed using an eye tracker which recorded the first movement of the eyes of the participant. This movement could be to either of two faces, effectively recording the natural, initial response. However, in this also lies the largest difference. As this research and the research of Stuit et al. (2021) both concern faces, one primarily focusses on initial attention drawn to a specific emotion/face and the other focusses more on the perception of ensembles, multiple faces. In the research regarding ensemble coding there is some controversiality regarding whether attention is required. Using divided attention tasks, it was found that ensemble statistics could be easily reported even without attention (Bronfman, Brezis, Jaconson & Usher, 2014; Alvarez & Olivia, 2008; Chong & Treisman, 2005b). However, there are also papers concluding attention facilitates ensemble perception. But what’s more is that ensemble perception seems to work most readily when attention cannot be deployed (McNair, Goodbourn, Shone & Harris, 2016; Oriet & Brand 2013; Joo, Chong & Blake, 2009). Therefore it seems that attention and ensemble coding are fundamentally different processes, with ensemble coding working quickly and with seamlessly little effort and attention requiring some form of cognitive strain or focus (Whitney & Yamanashi Leib, 2018). Based on the results from the current experiment it seems that while the distribution of attention is influenced by spatial frequency, high-level ensemble coding is not. However, there is still more research necessary to explore the connection between attention and ensemble perception.   
 To conclude, ensemble coding is especially useful at the highest levels of perceptual processing, it enables humans to quickly perceive characteristics of crowds or single out deviant emotions in a large group. Recent advancements in the field of face perception showed that spatial frequency greatly influences attention. However it seems that the perception of multiple faces is a different matter. The ensemble coding of emotions is most likely a primarily based on holistic stimuli. Further research into high-level ensembles could focus on decoding these holistic stimuli to provide better insights into the way we perceive the emotional states of large groups of people.

*Literature*

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

Alvarez GA, Oliva A. 2008. The representation of simple ensemble visual features outside the  
 focus of attention. Psychol. Sci. 19(4):392–98

Bronfman ZZ, Brezis N, Jacobson H, Usher M. 2014. We see more than we can report: “cost  
 free” color phenomenality outside focal attention. Psychol. Sci. 25(7):1394–403

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: salient physical  
 features guide effective visual search. *Journal of Experimental Psychology:  
 General*, *137*(3), 471.

Ceccarini, F., & Caudek, C. (2013). Anger superiority effect: The importance of dynamic  
 emotional facial expressions. *Visual Cognition*, *21*(4), 498-540.

Chong SC, Treisman A. 2005b. Statistical processing: computing the average size in perceptual  
 groups. Vis. Res. 45(7):891–900

Haberman J, Lee P, Whitney D. 2015b. Mixed emotions: sensitivity to facial variance in a  
 crowd of faces. J. Vis. 15(4):16

Haberman J, Whitney D. 2007. Rapid extraction of mean emotion and gender from sets of  
 faces. Curr. Biol. 17(7):751–53 www.annualreviews.org • Ensemble Perception 125  
 Annu. Rev. Psychol. 2018.69:105-129.

Haberman J, Whitney D. 2009. Seeing the mean: ensemble coding for sets of faces. J. Exp.  
 Psychol. Hum. Percept. Perform. 35(3):718–34

Haberman J, Whitney D. 2010. The visual system discounts emotional deviants when  
 extracting average expression. Atten. Percept. Psychophys. 72(7):1825–38

Han, L., Leib, A. Y., Chen, Z., & Whitney, D. (2021). Holistic ensemble  
 perception. *Attention, Perception, & Psychophysics*, *83*(3), 998-1013.

Hansen, C. H. & Hansen, R. D. Finding the face in the crowd: an anger superiority effect. *J.  
 Pers. Soc. Psychol.* **54**(6), 917 (1988).

Hodsoll, S., Viding, E., & Lavie, N. (2011). Attentional capture by irrelevant emotional  
 distractor faces. *Emotion*, *11*(2), 346.

Jeantet, C., Caharel, S., Schwan, R., Lighezzolo-Alnot, J. & Laprevote, V. Factors  
 influencing spatial frequencies extraction in faces: a review. *Neurosci. Biobehav.  
 Rev.* **93**, 123–138 (2018).

Joo SJ, Shin K, Chong SC, Blake R. 2009. On the nature of the stimulus information necessary  
 for estimating mean size of visual arrays. J. Vis. 9(9):7

Juth, P., Lundqvist, D., Karlsson, A., & Öhman, A. (2005). Looking for foes and friends:  
 perceptual and emotional factors when finding a face in the crowd. *Emotion*, *5*(4), 379.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency  
 processing during scene perception. *Frontiers in integrative neuroscience*, *8*, 37.

Kumar, D. & Srinivasan, N. Emotion perception is mediated by spatial frequency  
 content. *Emotion* **11**(5), 1144 (2011).

Landy MS. 2014. Texture analysis and perception. In The New Visual Neurosciences, ed. JS  
 Werner, LM Chalupa, pp. 639–52. Cambridge, MA: MIT Press

LoBue, V. (2009). More than just another face in the crowd: Superior detection of threatening  
 facial expressions in children and adults. *Developmental science*, *12*(2), 305-313.

Lundqvist, D., & Ohman, A. (2005). Emotion regulates attention: The relation between facial  
 configurations, facial emotion, and visual attention. *Visual Cognition*, *12*(1), 51-84.

McNair NA, Goodbourn PT, Shone LT, Harris IM. 2016. Summary statistics in the  
 attentional blink. Atten. Percept. Psychophys. 79:100–16

Oliva A, Torralba A. 2006. Building the gist of a scene: the role of global image features in  
 recognition. Prog. Brain Res. 155:23–36

Oriet C, Brand J. 2013. Size averaging of irrelevant stimuli cannot be prevented. Vis. Res.  
 79:8–16

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of experimental  
 psychology: human learning and memory*, *2*(5), 509.

Rensink, R. A., O'regan, 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.

Savage, R. & Lipp, O. V. The effect of face inversion on the detection of emotional faces in  
 visual search. *Cogn. Emot.* **29**(6), 972991.    
 <https://doi.org/10.1080/02699931.2014.958981> (2015).

Stuit, S.M., Kootstra, T.M., Terburg, D. *et al.* The image features of emotional faces that  
 predict the initial eye movement to a face. *Sci Rep* **11,**8287 (2021).  
 <https://doi.org/10.1038/s41598-021-87881-w>

Watamaniuk SN, McKee SP. 1998. Simultaneous encoding of direction at a local and global  
 scale. Percept. Psychophys. 60(2):191–200

Watamaniuk SN, Sekuler R, Williams DW. 1989. Direction perception in complex dynamic  
 displays: the integration of direction information. Vis. Res. 29(1):47–59

Whitney D, Haberman J, Sweeny T. 2014. From textures to crowds: multiple levels of  
 summary statistical perception. In The New Visual Neuroscience, ed. JS Werner, LM  
 Chalupa, pp. 695–710. Cambridge, MA: MIT Press

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual review of  
 psychology*, *69*, 105-129.