Different Roles of Foveal and Extrafoveal Vision in Ensemble Representation for Facial Expressions

Luyan Ji[1,2], Wenfeng Chen[1,*], and Xiaolan Fu[1]

[1]State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
{jily,chenwf,fuxl}@psych.ac.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, China

Abstract. People could extract mean expression of multiple faces pretty precisely. However, the mechanism of how we make such ensemble representation was far from clear. This study aimed to explore how faces in the foveal and extrafoveal vision contribute to the ensemble representation and whether the emotion of faces modulates the contribution. In the experiment, the expressions of foveal and extrafoveal faces were independently manipulated by changing the ratio of happy vs. angry faces. The participants reported whether the overall emotion was positive or negative. The results showed that faces in the foveal vision were given more weight than those in the extrafoveal vision in ensemble emotional representation. In addition, the ensemble perception was more accurate when faces in the extrafoveal vision were positive. These findings have great implications for the emotional design in interactive systems, especially when there are multiple users or multiple avatars presented on the screen.

Keywords: Ensemble representation, Facial expression, Foveal vision, Extrafoveal vision.

1 Introduction

Emotion is very important in our interaction with people as well as computers in everyday life. Recognizing users' emotions is the first step towards using emotions to improve adaptive performance of computers. Usually, the research on emotion recognition has been focused on a single person or an isolated face. However, in some occasions, like group studies or multiplayer games, there might be a number of people interacting with computer system at the same time. Nevertheless, the mechanism of how we acquire such ensemble representation was far from clear. For example, not all faces could fall into the foveal vision at a time, but little work concentrated on the role of locations in the visual field in ensemble perception. In this study, we aimed to investigate what the roles foveal faces and the extrafoveal faces play respectively in ensemble representation for facial expressions. It has great implications for the

* Corresponding author.

D. Harris (Ed.): EPCE 2014, LNAI 8532, pp. 164–173, 2014.

emotional design in interactive systems, especially when there are multiple users or multiple avatars presented on the screen.

Previous studies have demonstrated that observers can extract the mean emotion of multiple faces rather precisely [1-3]. This ability is robust and flexible, operating even on sets containing as many as 24 faces shown for only 100 ms [4]. However, when the set size is relatively large, observers cannot focus on all the faces simultaneously shown at one time. During normal viewing, our eyes direct a sequence of images to the fovea when sampling the visual scene and the fovea is a small region of the retina that corresponds to the central 2° of the visual field [5]. When the exposure time of face set was sufficient (e.g., a set of four faces shown for 2000 ms [2]), the observers could have time to freely scan every face; but if the set made up of large number of faces was shown very briefly [4], the observers could only glimpse a few faces or even some parts of a face. In the latter condition, when some faces or some features of one face occupy the fovea, on one fixation, other faces in the set would project extrafoveally.

The visual system indeed has the ability to detect or recognize emotional facial expressions at extrafoveal visual field, but the performance declined with increased eccentricity [6-8]. Neuroimaging study have shown that the affective modulation of early ERP components exists for both centrally and peripherally presented pictures, but the latter was lower in amplitude and slightly delayed [7]. The declined performance and neural response may be due to low visual acuity and low contrast sensitivity of extrafoveal vision [9]. The parvocelullar system which mainly begins in central part of the retina is particularly sensitive to contrast and high spatial frequency, like the detail of objects [10]; on the contrary, the magnocellular system which essentially originates from peripheral parts of the retina is less effective in contrast detection, but possesses a high temporal resolution [11-12]. According to these facts, it can be speculated that faces would be processed differently in the extrafoveal vision, contrast to in the foveal vision.

However, little is known about whether the foveated faces and the nonfoveated faces would contribute differently to ensemble representation for facial expressions. To our best knowledge, only one study paid attention to the relationship between eccentricity and the averaging of emotional expressions [13]. They asked the observers to judge the emotional intensity of the target face, which was shown isolated or surrounded by flankers, and the face (set) was presented either centrally or extrafoveally. The results revealed that judgment of the target face was less accurate and leaned more towards the average emotion of the flanker set when presented in the extrafoveal location. It seems to suggest that the impact of extrafoveal ensemble faces on individual recognition was greater than that of central ensemble faces. But it remains unclear as for the roles of the extrafoveal faces and foveal faces in ensemble representation. To tackle this issue, participants were instructed to report the ensemble emotion of a face set in the present study, instead of judging on one target face. We also developed a new display that both foveal and extrafoveal vision can be used, and that was different from gaze-contingent display [14-15], which restricted the observers to use central vision only or peripheral vision only. In addition, the central four faces would always occupy the foveal visual field by carefully manipulating the visual angle, and the remained surrounding faces would fall out of the fovea on one fixation. With this kind of display, the study could explore how observers represented the mean emotion when some faces were in foveal location while the other faces were in extrafoveal location.

Another purpose of this study was to investigate whether the role of the extrafoveal faces in the ensemble emotional representation would be enhanced or weakened by the emotional valence. It was evident that recognizing negative expressions (like anger, fear and sadness) is impaired when presented peripherally, whereas recognition of happiness suffered the least from peripheral presentation and was comparable with foveal performance [16]. However, controversy remains whether the extrafoveal vision favors for happy faces or for the negative faces. When an emotional face paired with a neutral face was shown in bilateral extrafoveal vision, happy face was identified faster than negative face indicated by shorter saccade latencies [17]; but the study adopting attention probe technique showed an advantage in rapid orienting of attention towards negative face, but not positive face [18]. There is also evidence suggesting that positive and negative faces showed similar efficiency in attentional capture, although both were more effective than neutral faces [19]. In general, it is far to reach a consistent conclusion on the role of extrafoveal facial emotions. Furthermore, few studies have focused on the attentional bias of ensemble representation for multiple facial expressions. So far, only one study found a positive bias of ensemble emotional representation, but did not explore whether ensemble emotional representation suffered from extrafoveal vision [4]. Therefore, the relationship between ensemble representation and extrafoveal emotion requires further investigation.

In the present study, we used happy and angry faces as emotion stimuli to investigate whether the foveal faces and the extrafoveal faces contribute differently in ensemble representation for multiple emotional expressions and whether there was an advantage for extrafoveal happy face or angry face in ensemble coding. Similar as Yang et al [4], we used face images from different people rather than morphed images of one person. We hypothesized that faces in the foveal vision, compared with faces falling out of the fovea, would be given higher weight in ensemble representation for facial expressions. Therefore, when the emotion of foveal faces was incongruent with that of extrafoveal faces, observers would judge the overall emotion based more on foveal faces. For the second question, there might be several possibilities: (a) The role of extrafoveal happy faces rather than angry faces was less impaired in the averaging of facial expressions, because the happy faces in the extrafoveal vision enjoyed an advantage in identification [16-17], (b) the role of extrafoveal angry faces rather than happy faces was less impaired in the ensemble representation for facial expressions, because the angry faces captured attention better than happy faces in the extrafoveal vision [18] , and (c) the role of extrafoveal happy and angry faces was impaired to the same extent, because the advantages were for both sides were comparable [19].

2 Method

2.1 Participants

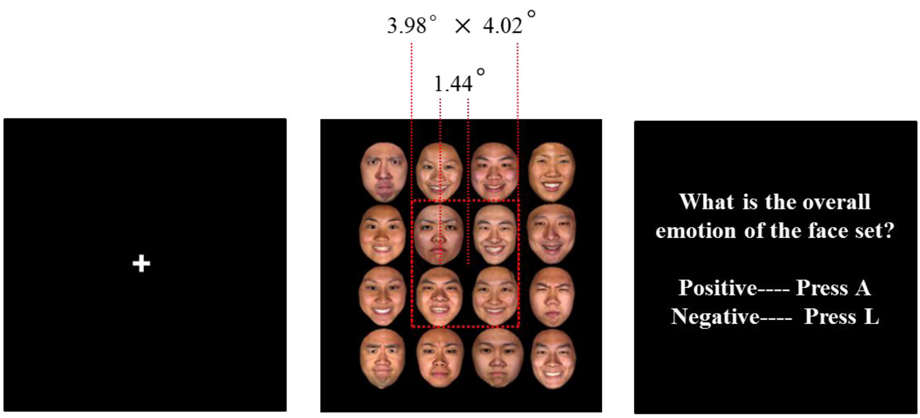
Forty-six volunteers (19-26 years old; 23 females) participated in the experiment. All of them had normal or corrected-to-normal vision and were paid for their time. Four participants did not properly respond to the face set (two responded too fast, and two too slow, which exceeded three standard deviations). Final analyses were based on the remaining 42 participants.

2.2 Stimuli

We selected 32 color photographs of faces from BU3DFE database [20], including 8 happy male faces, 8 happy female faces, 8 angry male faces and 8 angry female faces. All images were scaled to the same mean luminance and root-mean-square contrast [21]. Each face image subtended a visual angle of 1.97° × 1.97° at a viewing distance 85cm, and was presented against a black background.

Faces were presented in a 4 × 4 invisible grid, and the set of 16 face images occupied 6.40° × 6.53° of visual angle. For the central four faces, they occupied 3.98° × 4.02° of visual angle, falling into the foveal visual field (1.5mm diameter, about 4.3 degree [22]) and the eccentricity of each was 1.44° from the fixation point (Figure 1). The mean emotion of central four faces and that of surrounding twelve faces was independently manipulated by changing the ratio of happy to angry faces (Table 1). The ratio of happy vs. angry faces in the whole set was 3:1, 5:3, 3:5, or 1:3. Due to the number of surrounding faces outweighed that of central faces, the overall emotion of the whole set was always consistent with that of surrounding faces.

Face images in each set were randomly selected with two constraints: (1) an equal number of male and female models were presented in each set, and (2) no two images in each set were of the same model.



500ms 100ms Till response

Fig. 1. Experiment procedure and stimuli. In the sample face set, the mean emotion of central four faces is positive and that of surrounding faces is also positive. It is a congruent condition.

2.3 Design

The design involved two within-subject factors: congruency (congruent vs. incongruent) and extrafoveal emotion (positive vs. negative).

Table 1. Ratio of happy vs. angry faces in the foveal and extrafoveal vision respectively and the corresponding mean emotion of foveal and extrafoveal faces

Foveal H vs. A	Extrafoveal H vs. A	Foveal emotion	Extrafoveal emotion	Congruency[1]
4:0	8:4	Positive	Positive	C
4:0	2:10	Positive	Negative	IC
4:0	0:12	Positive	Negative	IC
3:1	9:3	Positive	Positive	C
3:1	7:5	Positive	Positive	C
3:1	3:9	Positive	Negative	IC
3:1	1:11	Positive	Negative	IC
1:3	11:1	Negative	Positive	IC
1:3	9:3	Negative	Positive	IC
1:3	5:7	Negative	Negative	C
1:3	3:9	Negative	Negative	C
0:4	12:0	Negative	Positive	IC
0:4	10:2	Negative	Positive	IC
0:4	4:8	Negative	Negative	C

2.4 Procedure

Each trial began with a 500ms fixation point at the center of the screen, followed by a set of faces presented for 100ms. Due to the short duration, participants were unlikely to execute a saccade (minimal saccade latency is about 150 ms [23]) to make the

[1] C denotes congruency and IC denotes incongruency. When the mean emotion of the foveal and extrafoveal faces are both positive or negative, it is the congruent condition. When the mean emotion of the foveal is positive whereas that of extrafoveal faces is negative, or vice versa, it is the incongruent condition.

surrounding 12 faces in the foveal vision. Immediately after the face set disappeared, participants were prompted to indicate whether the overall emotion of previous face set was positive or negative by pressing corresponding keys ("A" for positive and "L" for negative, vice versa).

Participants completed 280 trials, with 120 trials evenly divided into two congruent conditions and 160 trials evenly divided into incongruent conditions. There were 16 practice trials before the main experiment.

3 Results

3.1 Accuracy

The accuracies were calculated for each subject and each condition based on whether the subjective report conforms to the mean emotion of the whole set. A repeated-measure of ANOVA (congruency and extrafoveal emotion) was conducted. The results showed that the accuracy was significantly higher when the foveal and extra-foveal emotion were congruent than incongruent (F (1, 41) = 162.54, p<.001, η^2 = .80), and the accuracy was significantly higher when extrafoveal faces were happy than those were angry (F (1, 41) = 9.79, p<.01, η^2 = .19); but there was no interaction of congruency by extrafoveal emotion (F (1, 41) <1) (Figure 2). A t-test revealed that the accuracy was significantly lower than the chance level in the incongruent conditions (t (41) = -7.49, p< .001; t (41) = -5.92, p< .001).

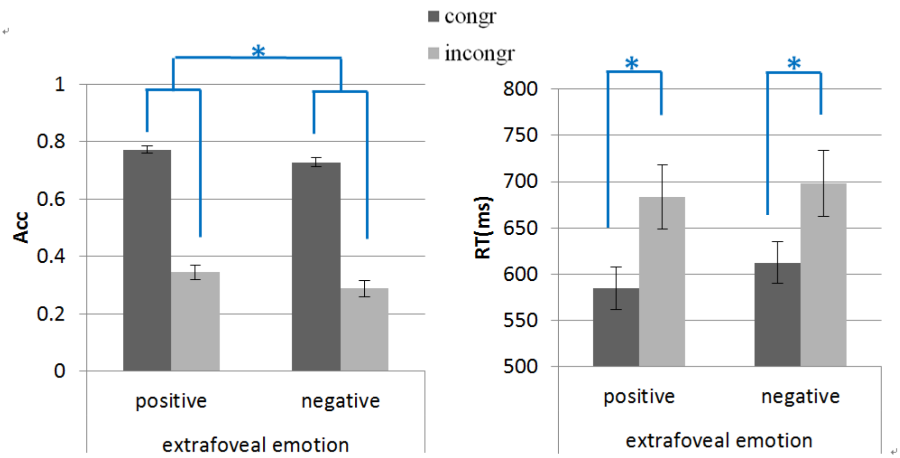


Fig. 2. Mean accuracies and RTs as a function of congruency and extrafoveal emotion. Error bars indicate ±1 standard error. Star marks indicate significant results.

3.2 Reaction Times

The reaction times for correct responses for each condition are shown in Figure 2. We conducted a 2 (congruency) by 2 (extrafoveal emotion) analysis of variance

(ANOVA) for repeated measures. The analysis revealed that the reaction times were faster when the foveal and extrafoveal emotion were congruent than incongruent (F (1, 41) = 24.70, p<.001, η^2 = .38); but there was no difference between extrafoveal happy face and extrafoveal angry face condition in reaction times (F (1, 41) = 3.10, p=.086, η^2 = .07) nor interaction between congruency by extrafoveal emotion (F (1, 41) <1).

4 Discussion

The present study showed that the foveal faces were given more weights than extrafoveal faces in ensemble representation. When the emotion of foveal faces and extrafoveal faces were incongruent, emotion in the foveal vision contributed more to the averaging, although the overall emotion was consistent with extrafoveal emotion rather than foveal emotion, so the accuracy was much lower and reaction times were much slower in incongruent conditions compared with congruent conditions. Besides, compared to extrafoveal angry faces, we were more sensitive to extrafoveal happy faces, and the happy expressions enhanced the role that extrafoveal faces played in ensemble representation.

4.1 Foveal Faces Contribute More to Averaging Facial Expressions

We found a below chance performance of ensemble representation when the foveal emotion and the extrafoveal emotion were incongruent, and indicated that subjects did not acquire a correct ensemble emotional representation of the face set. Since the overall emotion was always consistent with the extrafoveal emotion, the wrong response may suggest that subjects were prone to judge the overall emotion based on the foveal emotion, which was opposite to the extrafoveal emotion.

Previous research has found that not all items in a set contribute the same to the mean. The mean size estimations would bias towards the size of the attended item [24], and the emotional outliers in a set of faces would be discounted in the summary representation [25]. Besides, a recent eye movement study showed that the perceived ensemble expression is based on the particular faces that are fixated when the faces are randomly arranged [26]. In the present study, we are the first to explore the role of foveal and extrafoveal vision in ensemble representation for facial expressions. Consistent with hypothesis, the foveal faces played a much more important role in the ensemble representation compared with extrafoveal faces. The central four faces got the fixation and enjoyed the foveal vision; while because of limited duration time, the surrounding faces could not be fixed and were only exposed extrafoveally. The magnocellular system which essentially originates from peripheral parts of the retina was low in sensitivity to high spatial frequency [12], and removal of high spatial frequency would lead to great impairment in discriminating and recognizing facial expressions [16].

The faces in the extrafoveal vision are not only detected or recognized worse [6, 8], but are also down weighted in the ensemble representation, even though the number of extrafoveal faces are much larger. This result also suggests that the statistical properties of a set of items are not always extracted automatically, since the mean emotion of multiple faces should not be affected by foveal or extrafoveal vision. But there is a potential alternative that it is the difference of difficulty in foveal and extrafoveal processing that counts the different role of foveal and extrafoveal in the ensemble represenation, if the duration time was extended a little (but still within the minimal saccade latency) or the number of extrafoveal faces were reduced, observers might extract the mean emotion in a parallel way with distributed attention and then the different roles of foveal and extrafoveal vision might disappear. This alternative remains to be investigated in further research.

4.2 Advantage of Extrafoveal Happy Faces

Although the role of extrafoveal faces in the ensemble representation was impaired, the present study demonstrated that the role of extrafoveal happy faces was less impaired, compared with extrafoveal angry faces, no matter the congruency. So, the ensemble representation does not shut the door completely to the extrafoveal emotion.

The advantage for extrafoveal happy faces might due to physical feature saliency as well as positive affect. All the happy face images we used in the present study had a smile with an open mouth. It might be not necessary to recognize a happy face, but just detect the salient smile [16]. Especially when multiple happy faces were shown together, the open mouth with white teeth would be rather salient. On the other hand, some researchers pointed out that apart from featural processing, the affective processing plays a role that the priming effect of extrafoveal happy faces emerged at 750ms rather than 250ms [14]. In the present study, we are not aimed to explore whether the featural or affective processing contribute more to the advantage for extrafoveal happy faces, but it is worth further investigation and especially when there are multiple extrafoveal happy faces.

Then, why did not the angry faces enjoy an advantage in the extrafoveal vision? There are several possibilities. Firstly, the participants in the present study were generally in low level of anxiety and the angry faces might be only minor threatening. Different from the vigilance for severe threat [27], low vulnerable groups are prone to avoid the minor threat [28], so the processing of extrafoveal angry faces might be inhibited rather than strengthened. Secondly, the angry faces in the extrafoveal vision were not that salient compared with extrafoveal happy faces, and the physical saliency competed against the affective significance. Last but not least, extrafoveal angry faces hold attention and lead to a "disengagement deficit" from such stimuli [29], so processing extrafoveal angry faces might come at a cost. Future work will be needed to clear these possibilities and found out why there was an advantage for extrafoveal happy faces compared with angry faces. Maybe, we could use more threatening stimuli, like fearful faces, as well as anxious participants, to explore whether the advantage for extrafoveal happy faces still exist or whether there would be an advantage for extrafoveal negative faces instead.

4.3 Implications of the Present Study

The study found that foveal faces contribute more to the ensemble representation for facial expressions, and the extrafoveal happy faces were less impaired compared with extrafoveal angry faces. It also gives a new insight to the emotional design in interactive systems. For multiple users' interaction with computer, the computer has to combine the overall emotional information from all these users, but may give different weights to each individual. Users who stay closer to the screen may be more involved in interacting with the computer, so it makes sense for the computer to give more weights to the emotion expressed by these users when analyzing the overall emotional state of the users.

In addition, if there are multiple avatars showed on one screen, the designers should taker users' attentional capacity and visual field into consideration. Our visual system has limited attentional and short-term memory capacity [30-31]. If there is too much information on the computer screen, our brain may take use of the statistical regularity to condense the redundant information [32], but the extrafoveal information will be down-weighted or ignored to a large extent. So, the most important information or the key avatars should be arranged in the foveal or near the foveal vision of users.

On the other hand, the happy faces or other positive information shown extrafoveally play a more important role compared with extrafoveal negative information. When receiving a negative feedback from the computer, the users may feel better if there was also something positive, even if not presented at the corner of the screen.

Acknowledgement. This research was supported in part by grants from National Basic Research Program (2011CB302201), and the National Natural Science Foundation of China (31371031).

References

1. Haberman, J., Harp, T., Whitney, D.: Averaging Facial Expression over Time. Journal of Vision 9(11), 1–13 (2009)
2. Haberman, J., Whitney, D.: Rapid Extraction of Mean Emotion and Gender from Sets of Faces. Current Biology 17(17), R751–R753 (2007)
3. Haberman, J., Whitney, D.: Seeing the Mean: Ensemble Coding for Sets of Faces. Journal of Experimental Psychology: Human Perception and Performance 35(3), 718–734 (2009)
4. Yang, J.W., Yoon, K.L., Chong, S.C., Oh, K.J.: Accurate but Pathological: Social Anxiety and Ensemble Coding of Emotion. Cognitive Therapy and Research 37, 572–578 (2013)
5. Atkinson, A.P., Smithson, H.E.: Distinct Contributions to Facial Emotion Perception of Foveated versus Nonfoveated Facial Features. Emotion Review 5(1), 30–35 (2013)
6. Bayle, D.J., Schoendorff, B., Henaff, M.A., Krolak-Salmon, P.: Emotion Facial Detection in the Peripheral Visual Field. PLoS ONE 6(6), e21584 (2011)
7. Rigulot, S., Delplanque, S., Despretz, P., Defoort-Dhellemmes, S., Honore, J., Sequeira, H.: A Spatiotemporal Analysis of Early ERP Responses. Brain Topography 20, 216–223 (2008)
8. Rigulot, S., D'Hondt, F., Honore, J., Sequeira, H.: Implicit Emotional Processing in Peripheral Vision: Behavioral and Neural Evidence. Neuropsychologia 50, 2887–2896 (2012)
9. Robson, J.G., Graham, N.: Probability Summation and Regional Variation in Contrast Sensitivity across the Visual Field. Vision Research 21, 409–418 (1981)
10. Dacey, D.M., Petersen, M.R.: Dendritic Field Size and Morphology of Midget and Parasol Ganglion Cells of the Human Retina. Proceedings of the National Academy of Sciences 89, 9666–9670 (1992)
11. Livingston, M.S., Hubel, D.H.: Psychophysical Evidence for Separate Channels for the Prception of Form, Color, Movement, and Depth. Journal of Neuroscience 7, 3416–3468 (1987)

12. Schiller, P.H., Logothetis, N.K., Charles, E.R.: Functions of The Colour-opponent and Broad-band Channels of the Visual System. Nature 343, 68–70 (1990)
13. Fielding, K.M., Carvey, R.J., Liu, C.H.: Singular or Summary: Averaging of Facial Expression in Sets is Modulated by Eccentricity. Journal of Vision 13(9), 588 (2013)
14. Calvo, M.G., Nummenmaa, L., Avero, P.: Recognition Advantage of Happy Faces in Extrafoveal Vision: Featural and Affective Processing. Visual Cognition 18, 1274–1297 (2010)
15. Reingold, E.M., Loschky, L.C., McConkie, G.W., Stampe, D.M.: Gaze-contingent Multiresolutional Displays: An Integrative Review. Human Factors 45, 307–328 (2003)
16. Goren, D., Wilson, H.R.: Quantifying Facial Expression Recognition across Viewing Conditions. Vision Research 46, 1253–1262 (2006)
17. Calvo, M.G., Nummenmaa, L.: Eye-movement Assessment of the Time Course in Facial Expression Recognition: Neurophysiological Implications. Cognitive, Affective & Behavioral Neuroscience 9(4), 398–411 (2009)
18. Pourtois, G., Grandjean, D., Sander, D., Vuilleumier, P.: Electrophysiological Correlates of Rapid Spatial Orienting towards Fearful Faces. Cerebral Cortex 14, 619–633 (2004)
19. Vuilleumier, P., Schwartz, S.: Emotional Facial Expressions Capture Attention. Neurology 56, 153–158 (2001)
20. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: A 3D facial Expression Database for Facial Behavior Research. In: 7th IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2006), pp. 211–216. IEEE Press, Southampton (2006)
21. Bex, P.J., Makous, W.: Spatial Frequency, Phase, and the Contrast of Natural Images. Journal of the Optical Society of America 19, 1096–1106 (2002)
22. Riordan-Eva, P.: Anatomy & Embryology of the Eye. In: Riordan-Eva, P., Cunningham, E.T. (eds.) Vaughan & Asbury's General Ophthalmology, 18th edn., p. 13. McGraw-Hill Medical (2011)
23. Rayner, K.: Eye Movements in Reading and Information Processing: 20Years of Research. Psychological Bulletin 124, 372–422 (1998)
24. de Fockert, J.W., Marchant, A.P.: Attention Modulates Set Representation by Statistical Properties. Perception & Psychophysics 70(5), 789–794 (2008)
25. Haberman, J., Whitney, D.: The Visual System Discounts Emotional Deviants When Extracting Average Expression. Attention, Perception, & Psychophysics 72(7), 1825–1838 (2010)
26. Wolfe, B., Kosovicheva, A.A., Leib, A.Y., Whitney, D.: Beyond Fixation: Ensemble Coding and Eye Movements. Journal of Vision 13(9), 710 (2013)
27. Koster, E.H.W., Crombez, G., Verschuere, B., DeHouwer, J.: Selective Attention to Threat in the Dot Probe Paradigm: Differentiating Vigilance and Difficulty to Disengage. Behaviour Research and Therapy 42(10), 1183–1192 (2004)
28. Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M.J., van IJzendoorn, M.H.: Threat-related Attentional Bias in Anxious and Nonanxious Individuals: A Meta-Analytic Study. Psychological Bulletin 133(1), 1–24 (2004)
29. Fox, E., Russo, R., Bowles, R., Dutton, K.: Do Threatening Stimuli Draw or Hold Visual Attention in Subclinical Anxiety? Journal of Experimental Psychology: General 130(4), 681 (2001)
30. Luck, S.J., Vogel, E.K.: The Capacity of Visual Working Memory for Features and Conjunctions. Nature 390(6657), 279–281 (1997)
31. Scholl, B.J., Pylyshyn, Z.W.: Tracking Multiple Items through Occlusion: Clues to Visual Objecthood. Cognitive Psychology 38(2), 259–290 (1999)
32. Haberman, J., Whitney, D.: Ensemble Perception: Summarizing the Scene and Broadening the Limits of Visual Processing. Chapter to Appear in an Edited Volume. A Festschrift in Honor of Anne Treisman (2011)