Psychological Research
DOI 10.1007/s00426-014-0548-6

ORIGINAL ARTICLE

Local feature suppression effect in face and non-face stimuli

**Artyom Zinchenko · Hyojung Kim ·
Adrian Danek · Hermann J. Müller ·
Dragan Rangelov**

Received: 24 July 2013 / Accepted: 7 February 2014

Abstract There is evidence that the cognitive system processes human faces faster and more precisely than other stimuli. Also, faces summon visual attention in an automatic manner, as evidenced by efficient, 'pop-out' search for face targets amongst homogeneous non-face distractors. Pop-out for faces implies that faces are processed as a basic visual 'feature' by specialized face-tuned detectors, similar to the coding of other features (e.g., color, orientation, motion, etc.). However, it is unclear whether such face detectors encode only the global face configuration or both global and local face features. If the former were correct, the face detectors should be unable to support search for a local face feature, rendering search slower relative to non-face stimuli; that is, there would be local feature suppression (LFS) for faces. If the latter was the case, there should be no difference in the processing of local and, respectively, global face features. In two experiments, participants discerned the presence (vs. absence) of a local target defined as a part of either a normal or a scrambled (schematic or realistic) face or of a non-face (Kanizsa diamond or realistic house) configuration. The results consistently showed a robust LFS effect in both reaction times and error rates for face stimuli, and either no difference or even a local feature enhancement effect for the control stimuli. Taken together, these findings indicate that faces are encoded as a basic visual feature by means of globally tuned face detectors.

A. Zinchenko (✉) · H. Kim · H. J. Müller · D. Rangelov (✉)
Allgemeine und Experimentelle Psychologie, Ludwig-Maximilians-Universität, 80802 Munich, Germany
e-mail: zinchenko@cbs.mpg.de

D. Rangelov
e-mail: Rangelov@psy.lmu.de

A. Danek
Neurologische Klinik, Ludwig-Maximilians-Universität, Munich, Germany

H. J. Müller
Department of Psychological Sciences, Birkbeck College, University of London, London, UK

Introduction

Perception of faces plays an important role in the life of every human being. It provides people with valuable information about the gender, age, verbal gestures, and emotional states of individuals around them. Faces are salient stimuli and capable of capturing people's attention more easily than other stimuli. For example, humans need only 100 ms to make a saccade toward a face, relative to 140 ms for a saccade to a non-face (e.g., vehicle) stimulus (Crouzet, Kirchner, & Thorpe, 2010); in addition, only 360 ms are necessary to correctly decide whether or not a face is familiar (Barragan-Jason, Besson, Ceccaldi, & Barbeau, 2013), suggesting that detecting and identifying a face are very fast processes. Furthermore, studies showed that people require <100 ms to detect a change in the face (Carbon, 2011) and only 100 ms to judge attractiveness, likeability, and trustworthiness of the presented face (Willis & Todorov, 2006). At the same time, it is harder to orient one's attention away from neutral (Bindemann, Burton, Hooge, Jenkins, & de Haan, 2005) or angry faces (Devue, Belopolsky, & Theeuwes, 2011), relative to other stimuli. The visual preference for faces appears very early in the childhood: newborns spend more time looking at normal compared to scrambled and inverted faces, as well as other non-face stimuli (Batki, Baron-Cohen, Wheelwright, Connellan, & Ahluwalia, 2000). Taken together, the available evidence suggests that faces are processed differently to other stimuli. The present study was designed to contribute to clarifying the processes underlying these differences.

The evidence that faces are processed faster than non-face stimuli raises the question as to the stage of processing—pre-attentive and/or post-selective—that is influenced by having to process faces. On the one hand, encountering faces in the visual scene might influence coding processes prior to the allocation of focal attention. Alternatively, the speed of spatial-attentional selection may be comparable for face and non-face stimuli, and having to process faces may rather influence post-selective perceptual analysis. The former, pre-attentive bias for face (relative to other) stimuli would imply that face-like stimuli in the visual scene are assigned higher attentional priority and are, thus, the first stimuli to be selected for focal-attentional processing. In other words, properties of face-like stimuli might guide visual selection. The latter—that is, post-selective—face processing bias would imply that face-specific processes come into play only after a stimulus has been attentionally selected. Reasons for this might be the inherent social relevance of faces that engenders more detailed analysis of face stimuli relative to other stimulus types.

If face properties were to guide attentional selection such that faces would be selected faster than other types of stimuli, then a face 'target' should always be the first attended stimulus in a scene consisting, besides the target, of other types of 'distractor' stimuli, and the speed of selecting the target should not vary as a function of the number of distractors. In other words, preferential pre-attentive processing of faces should yield efficient visual search or 'pop-out' for face targets. Several previous studies tested attentional bias for faces using schematic drawings of faces (Jolicoeur, 1994; Nothdurft, 1993), or black-and-white photographs (Brown, Huey, & Findlay, 1997), creating an odd-one-out target face and homogeneous distractor (inverted) faces, and manipulating the number of items in the search display (the 'set size'). Against the expectations, the results of these studies revealed mean search reaction times (RTs) to increase proportional to set size (Jolicoeur, 1994), and eye movement analyses did not reveal specific processing for faces (Brown, Huey, & Findlay, 1997); that is: faces did not pop out of the display. Importantly, Jolicoeur (1994) also reported that search became more efficient with distractors less similar to a face. Taken together, studies using a visual search paradigm with a particular face target among distractors that resembled a face consistently failed to show pop-out for face targets. However, by virtue of using face-like distractors, these studies only showed that local face features fail to guide attention, but not that global face properties (i.e., stimulus 'faceness') are unable to guide attention efficiently.

More recently, Hershler and Hochstein (2005, 2006) tested the ability of global face properties to guide attention, presenting search arrays consisting of a face amongst a heterogeneous set of realistic non-face stimuli (e.g., car, flower, glove, etc.). The results of these studies showed that faces do pop out when presented against non-face-like stimuli as distractors (however, see VanRullen, 2006; Kietzmann & Konig, 2010, for different results and conclusions). Importantly, non-face stimuli (e.g., a car) did not pop out with faces as distractors, suggesting that faces and non-faces are selected in a different manner. Moreover, since Hershler and Hochstein (2005, 2006) used many different distractors (e.g., car, flower, glove, etc.) that varied in color, shape, and orientation, it was unlikely that participants used a single, basic-level feature (e.g., shape, color, etc.) to discern the target from the distractors. Given this, Hershler and Hochstein (2005, 2006) hypothesized that efficient search for faces implies that they are detected on the basis of global, rather than local, stimulus properties and that faces, although composite stimuli, are processed as a single feature.

Evidence that faces pop out against non-faces suggests that they are processed similar to other basic visual features, such as color or orientation, for which pop-out effects have also been demonstrated (e.g., Foster & Ward, 1991; Treisman & Gormican, 1988; Wolfe, O'Neill, & Bennett, 1998). Pop-out effects for basic visual features are typically explained by assuming the existence of special, feature-specific detectors that, when activated, signal the presence of a particular feature (e.g., the color red) in the visual field (Cave & Wolfe, 1990; Itti, Koch, & Niebur, 1998). Dominant theories of attentional selection (e.g., Guided-search, Wolfe, Cave, & Franzel, 1989) and computational implementations (Itti, 2005) show that for pop-out, *two* conditions must be met: (1) feature detectors tuned to particular stimulus properties should exist, and (2) the output of these feature detectors should be spatially specific. To illustrate, presenting a single face among non-faces would result in face detectors signaling face presence at a single location, permitting reliable attentional allocation to that location. Within this framework, empirical evidence that faces pop out of the scene would imply the existence of face detectors that analyze the field in terms of the presence (vs. absence) of faces across different locations.

An important question arising from the findings of Hershler and Hochstein (2005, 2006) is: what do face detectors actually detect? On the one hand, these detectors might be tuned to a particular combination of local facial features (e.g., nose, mouth, eyes, etc.), simultaneously encoding the values of these features (e.g., nose width, mouth size, etc.) and their spatial arrangement.

Alternatively, face detectors might be tuned to the global face configuration (i.e., two blobs for eyes, a vertical line for a nose, a horizontal line for a mouth, and an oval head contour), without encoding precise local face properties. Since all human faces share the global face properties, while individual faces differ in their local features, the question about the nature of the face detectors can be re-formulated in terms of whether face detectors encode both the 'faceness' and (individual) identity of a face or the 'faceness' only. The existence of face detectors simultaneously tuned to both global and local face properties would predict superior processing of local as well as global face attributes, relative to attributes of stimuli for which such detectors are unlikely to exist.

Empirical findings appear to support superior processing of local face features (e.g., Murray, 2002; Rouw & Gelder, 2002; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). For instance, Rouw and Gelder (2002) presented participants with sets of either normal or scrambled faces. Each trial display consisted of a standard, or 'template', face and two comparison faces, with all three faces being either normal or scrambled. One of the comparison faces differed from the template in either the mouth or the eye features, while the second face was identical to the template. The task was to indicate which of the comparison faces was identical to the template. Rouw and Gelder (2002) observed faster RTs for normal as compared to scrambled faces, that is, the processing of local features was enhanced for normal faces [local feature enhancement (LFE), effect]. This finding points to facilitated processing of local facial features, which in turn implies that face detectors are tuned to local face components (e.g., nose, mouth, etc.).

Importantly, studies demonstrating superior encoding of local face features primarily used discrimination tasks; that is: to solve the task, participants had to encode local face features (see also Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Lobmaier, Bölte, Mast, & Dobel, 2010; Schwaninger, Lobmaier, & Collishaw, 2002; Leder & Carbon, 2005). By contrast, discerning the presence (vs. absence) of a face amongst non-faces might be achieved by global face detectors; that is, to reliably detect a face, it would in principle be sufficient to detect a stimulus displaying 'faceness', without necessarily encoding its identity. In this respect, Hershler and Hochstein (2006) argued for a dissociation between processes involved in face search and, respectively, processes involved in discrimination tasks—with search processes relying on face detectors that encode global face properties, as opposed to identification processes which rely on different representational units encoding local face properties precisely. Some findings show that while face inversion influences both face detection and identification, it has a much greater influence on the latter (Valentine, 1988; Yin, 1969). This suggests that face detectors encode both normal (upright) and inverted (upside-down) faces, and post-selective face identification relies on orientation-specific information.

The aim of the study

The present study was designed to investigate what kind of information is encoded by the hypothesized face detectors: (1) only global face attributes or (2) both global and local features. If the face detectors encoded only global properties, then presenting face stimuli that differ only in their local elements would activate the detectors in the same way whether the faces do or do not contain a local target. Consequently, signals coming from face detectors would fail to guide attention to the target location. However, the fact that face detectors are active would increase the overall activation for target and distractor configurations alike, rendering the local target less conspicuous relative to scrambled stimuli which do not activate detectors tuned to global configurations. Thus, on the assumption of globally tuned face detectors, worse performance would be expected when the target belongs to a normal face as compared to a scrambled configuration, that is: there would be a local feature suppression (LFS) effect. If faces really constitute a special class of stimuli, faces should be processed differently to other types of composite stimuli.

To preview the results, using schematic and realistic stimuli, as well as different local face properties as targets, all three experiments presented below revealed substantial LFS for faces exclusively—supporting the idea that face detectors are indeed preferentially tuned to global face properties, or stimulus "faceness". Alternative mechanisms that could give rise to the LFS for faces are considered in more detail in "General discussion".

Experiment 1

Experiment 1 consisted of two separate parts (Experiments 1a and 1b). In both parts, schematic line drawings were used as stimulus material. However, while Experiment 1a used more realistic face proportions (adapted from Powell & Humphreys, 1984) but less controlled local stimulus properties (i.e., number of local features within a face and non-face stimuli differed), Experiment 1b implemented less realistic proportions, but better local control.

Method

Participants

Twenty observers (4 males, mean age 25 years) took part in Experiment 1a, and a different group of 20 participants

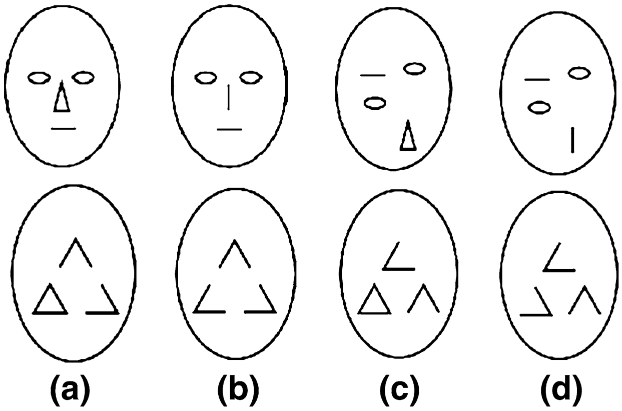


Fig. 1 Illustrates normal (**a**, **b**) and scrambled (**c**, **d**) face and Kanizsa stimuli used in Experiment 1a. **a, c** Target-present conditions, and **b, d** target-absent conditions, for both face (*upper*) and Kanizsa (*lower row*) stimuli

(2 males, mean age 24 years) took part in Experiment 1b. All participants had normal or corrected-to-normal vision and were naïve with respect to the purpose of the experiment. All participants gave informed consent prior to performing the task.

Apparatus

The experiments were controlled by a Dell PC running under the Windows XP operating system. Stimuli were presented on a 19″ CRT monitor, with a screen resolution of 1,024 × 768 pixels and a refresh rate of 85 Hz. The experimental software was custom written in PsychoPy (Peirce, 2007, 2009). Head-to-monitor distance was 53 cm, controlled by means of a chin rest. Participants responded by pressing the left or right mouse button with their left- or right-hand index finger, respectively. Response-assigned buttons were counterbalanced across participants.

Stimuli

In Experiment 1a, two types of stimuli were presented in two separate experimental sessions: (1) schematic faces (see Fig. 1a, b, upper row), and (2) non-face-like stimuli—Kanizsa triangles (Fig. 1a, b, lower row). Elements of both faces and Kanizsa configurations were presented, in separate blocks of trials, either (1) in their normal/canonical or (2) in scrambled configurations (Fig. 1c, d, both rows for faces and triangles, respectively). Schematic faces were used, rather than realistic face pictures, because they allowed for a better match between face and non-face stimuli. To make a schematic face look as close as possible to the realistic face, we used proportions adapted from Powell and Humphreys (1984). In this way, the schematic faces closely resembled realistic faces by maintaining the correct location of the face parts (eyes, nose, and mouth) in relation to each other. Normal faces consisted of an oval with several geometrical elements inside that, together, composed a face in a neutral emotional state (Fig. 1a, b, upper row).

The scrambled faces consisted of the same set of elements (Fig. 1c, d, upper row); however, the location of each particular element was changed such that the global configuration no longer resembled a face. To increase the generalizability of the results for scrambled stimuli, three different configurations were constructed, and the analyses were performed over averages across the three. The scrambled faces were used to keep the target stimuli (a triangle in Experiments 1a and 1b and mouth in Experiment 2) identical across the control and experimental conditions. Applying another commonly used procedure to disrupt global face processing such as inversion would have also rendered the target stimuli physically different between conditions.[1] Consequently, it would remain equivocal whether our (anticipated) effect pattern was indeed attributable to differences in configurations, or simply to physical target differences.

A similar procedure was adopted to create non-face stimuli. Different line elements were used to create a global triangle shape (Fig. 1a, b, lower row). To make these stimuli comparable to those used in the face condition, Kanizsa triangles were placed inside the ovals. The scrambled Kanizsa stimuli (of which there were three different configurations) were composed of the same constituents as in normal Kanizsa triangles, but these were rotated differently so as not to form a unitary ‘Gestalt’ (Fig. 1c, d, lower row).

In Experiment 1b, Kanizsa stimuli consisted of four triangular elements arranged so as to form a global diamond shape, thus matching the number of local elements in the face stimuli. Also, physically identical triangles served as the targets in both the face and the Kanizsa stimuli (Fig. 2). The stimulus eccentricity was somewhat reduced to 5.82° of visual angle, making the stimuli easier to process. The size of each stimulus oval was 2.96° × 4.16°, and the size of the target triangle was 0.28° × 0.92° of visual angle.

Six oval stimuli (subtending area of 4.1° × 2.76° of visual angle) were equidistantly distributed around the

---

[1] Furthermore, since previous studies (e.g., Kanwisher, Tong, & Nakayama, 1998; Murray, 2002; Richler, Mack, Palmeri, & Gauthier, 2011), showed that some components of face processing are relatively unaffected by inversion using inverted faces would have rendered our paradigm less sensitive. For example, Kanwisher et al. (1998) showed that inverted faces still activate face-specific FFA, while non-face objects (in whatever orientation) do not, suggesting that inverted faces still engage face-specific brain mechanisms.

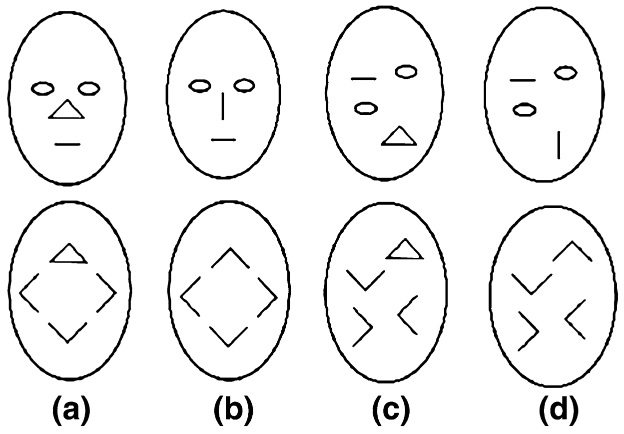


Fig. 2 Illustrates normal (**a**, **b**) and scrambled (**c**, **d**) face and Kanizsa stimuli used in Experiment 1b. **a, c** Target-present conditions, and **b, d** target-absent conditions, for both face (*upper*) and Kanizsa-type stimuli (*lower row*)

circumference of a virtual circle, of a radius of 7.15° of visual angle. The target stimulus was a triangle located within an oval. For faces, the target triangle subtended an area of 0.41° × 0.72°, while the target for Kanizsa stimuli subtended 0.28° × 0.92° of visual angle. When present, the target was placed randomly in one of the six ovals in the display. Local features of ovals were made task relevant, while the overall global configuration and semantic meaning of stimuli was absolutely task irrelevant.

The distractors were similar to the target configurations, that is, the six ovals making up the display were all the same except for one oval that could contain the local target feature (on target-present trials). Had we used heterogeneous distractors, our participants would have been able to detect the target on the basis of its local (i.e., triangle singleton) as well as its global (i.e., faceness) properties. Since our study was designed to examine how the local properties of a complex stimulus are processed, using homogeneous distractors in the present experiment was essential to prevent potential attentional selection on the basis of global stimulus properties.

Procedure

Participants were seated in a dimly lit and acoustically isolated room in front of the computer monitor. They were explicitly instructed to maintain fixation to the fixation cross (in the screen center) throughout the trial. To control for eye movements, the stimuli were presented for 200 ms only and then removed. The task was to discern the presence (70 % of trials) vs. absence (30 %) of the target among the distractor items as fast and accurately as possible. Error responses were followed by an 'Error' feedback message. To enforce fast reaction times, if no response was made within 1,500 ms, "Try to respond faster" feedback was presented, and such trials were excluded from further analysis. The sequence of events on a typical trial is illustrated in Fig. 3.

The experimental conditions were split into two sessions (i.e., face and Kanizsa). Each session consisted of 6 blocks (i.e., three scrambled and three normal blocks) with 50 trials per block, alternating between normal and scrambled configurations with the exact scrambled configuration being fixed per block (e.g., only S1). The order of session and stimulus type presentation was counterbalanced across participants.

Data treatment

The trials were first sorted according to (1) target type (present/absent), (2) stimulus type (faces vs. Kanizsa) and (3) stimulus configuration (normal vs. scrambled), yielding eight experimental conditions. Response times more than ±2 standard deviations from the individual mean RT per condition were considered as outliers (2 % of data) and excluded from further analysis. Mean RTs for correct responses and error rates per participant per condition were computed and submitted to repeated-measures ANOVAs with main terms for target type, stimulus type and stimulus configuration, and effects with $\alpha < 0.05$ considered significant. Mean RTs and error rates across different experimental conditions are depicted in Figs. 4 and 5.

Results

Inspection of the mean RTs across target type (present/absent), stimulus type (faces/Kanizsa) and configuration (normal/scrambled) in Experiment 1a (Fig. 4) showed RTs to be faster on target-present than on target-absent trials (523 vs. 676 ms, $p < 0.01$) and faster to face than to Kanizsa stimuli (579 vs. 621 ms, $p < 0.04$), with no difference between normal and scrambled stimuli (599 vs. 601 ms, $p = 0.76$). The three-way interaction between stimulus type, target type and configuration was also significant [$F(1,19) = 4.83$, $\eta_p^2 = 0.20$, $p < 0.05$], attributable to a 33-ms LFS effect (i.e., normal–scrambled) for target-present trials for faces ($p < 0.01$) and no significant effect in the corresponding condition for Kanizsa stimuli (−16 ms, $p > 0.05$). For target-absent trials, the configuration effect was significant neither for faces (7 ms, $p > 0.05$), nor Kanizsa stimuli (7 ms, $p > 0.05$) indicative of the LFS effect being specific to target-present trials.

Inspection of the errors across target types, stimulus types, and stimulus configurations showed marked overall differences between target-present (miss errors) and target-absent trials (false-alarm errors) (15 vs. 44 %, $p < 0.01$);

Fig. 3 Illustrates the sequence of events on a trial in Experiments 1a and 1b. The fixation cross, presented for 1,000 ms, was followed by the visual search display, presented for 200 ms. The response had to be given within a 1,500-ms time limit

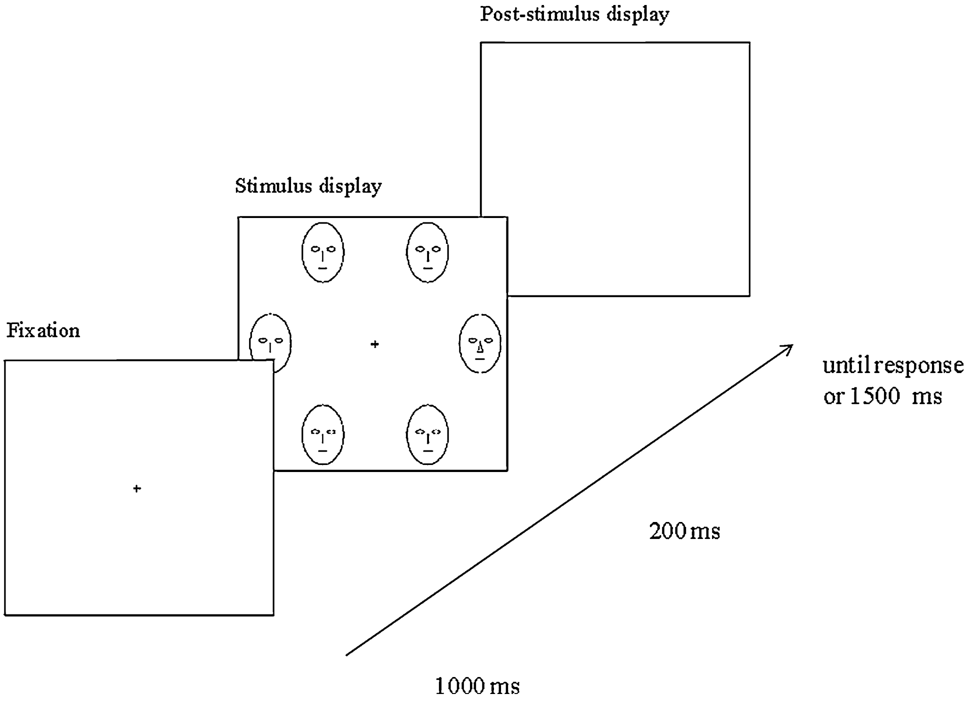
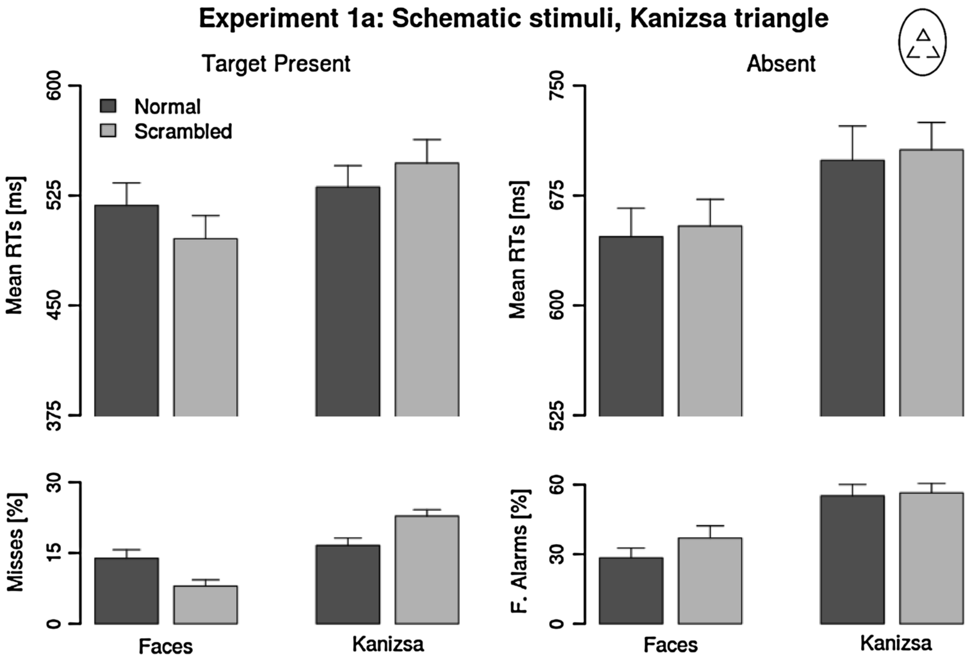


Fig. 4 Reaction times and error rates across experimental conditions of Experiment 1a. *Whiskers* denote 1 SEM. Notice the difference in scales for target-present (*left panel*) and -absent conditions (*right*)



between faces and Kanizsa stimuli (20 vs. 38 %, $p < 0.01$); and between normal and scrambled configurations (28 vs. 31 %, $p < 0.05$). Furthermore, the three-way interaction between stimulus type, target type and configuration was also significant [$F(1,19) = 10.34$, $\eta_p^2 = 0.35$, $p < 0.01$]: while for target-present trials there were significant configuration effects for both faces (6 % LFS, $p < 0.01$) and Kanizsas (6 % LFE, $p < 0.01$), the respective effects failed to reach significance for target-absent trials.

Figure 5 presents the mean RTs and error rates for the target-present trials of Experiment 1b. Overall, mean RTs were slower on target-absent than on -present trials (598 vs. 473 ms, $p < 0.01$). Furthermore, RTs were slower for normal than for scrambled faces (546 vs. 511 ms, i.e., 35 ms LFS, $p < 0.01$), and faster for normal than for scrambled Kanizsa stimuli (532 vs. 553, i.e., 21 ms LFE, $p < 0.01$), as confirmed by a significant stimulus type x configuration interaction [$F(1,19) = 12.77$, $\eta_p^2 = 0.40$,

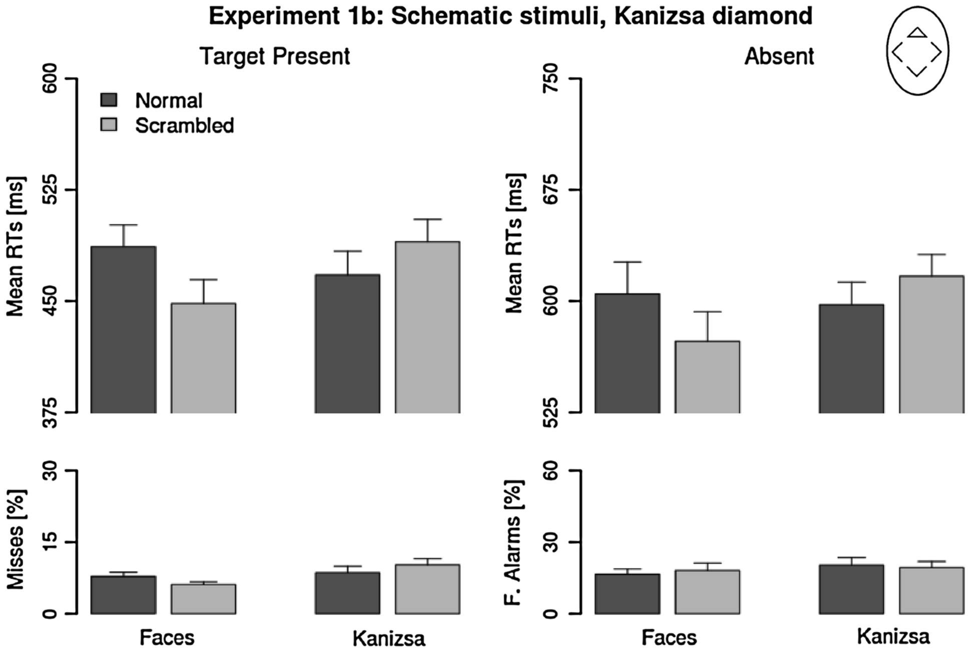


Fig. 5 Reaction times and error rates across experimental conditions of Experiment 1b. *Whiskers* denote 1 SEM. Notice the difference in scales for target-present (*left panel*) and -absent conditions (*right*)

$p < 0.01$]. No other main effects or interactions reached significance. In contrast to Experiment 1a, the LFS effect for faces was significant for both target-present (39 ms LFS, $p < 0.01$) and target-absent trials (32 ms LFS, $p < 0.01$). For Kanizsa stimuli, significant LFE was observed on target-present trials (23 ms LFE, $p < 0.05$), and a similar, albeit insignificant, numerical trend on target-absent trials (20 ms LFE, $p > 0.05$). Finally, analyses of error rates across conditions of Experiment 1b proved only the main effect of target type significant [$F(1,19) = 42.8, \eta_p^2 = 0.69, p < 0.01$] with more errors for target-absent than -present trials (19 and 8 %, respectively).

Discussion

The results of both Experiments 1a and 1b showed that face and non-face stimuli are processed differently. There was significant LFS for face stimuli in both mean RTs and error rates, suggestive of faces being processed as a single feature, with less-efficient processing of the local features of a face relative to a scrambled stimulus. The opposite was true for the Kanizsa stimuli: normal Kanizsa stimuli were associated with fewer errors and faster RTs compared to scrambled stimuli. These results are consistent with the hypothesis that face detectors encode the global face configuration; that is, faces are detected as a unique, global 'feature', rendering perception of local face features less efficient (i.e., LFS) relative to scrambled faces. This LFS effect is striking when contrasted with Kanizsa stimuli, which exhibited LFE (rather than suppression).

Importantly, the results of both Experiments 1a and 1b also revealed significant differences between the miss and false-alarm rates, with twice as many false alarms as misses. This suggests that participants adopted a liberal response criterion, which may not be surprising given the 2:1 ratio of target-present to target-absent trials. However, while a similar liberal response criterion was adopted with both face and Kanizsa configurations, LFS effects were observed exclusively for faces—arguing against the LFS effect being simply attributable to the choice of response criterion. Furthermore, Experiments 1a and 1b strongly differed in task difficulty, especially for Kanizsa stimuli, with Kanizsa triangles (Experiment 1a) resulting in much higher error rates than Kanizsa diamonds (Experiment 1b). However, task difficulty, too, could not readily account for the LFS effects for faces, which were evident both when the task was more (Experiment 1a) and when it was less difficult (1b). Thus, while response criterion and task difficulty strongly influenced performance in Experiments 1a and 1b, the LFS effects observed are likely to reflect properties of the perceptual processing of faces (rather than reflecting non-perceptual processes related to response selection).

Experiment 2

Experiment 2 was designed to investigate whether or not the results observed for schematic faces in Experiment 1 would generalize to more realistic face stimuli. Accordingly, Experiment 2 was similar in design to Experiment 1, except that schematic faces and Kanizsa stimuli

(Experiment 1) were replaced by more realistic pictures of faces and houses (Experiment 2). The use of more realistic stimuli was also meant to control for potential differences in familiarity between faces and Kanizsa stimuli, which might have influenced the results of Experiment 1.

Method

Participants

Twenty students (3 males, mean age 24 years) participated in Experiment 2. All participants had normal or corrected-to-normal vision and were naïve with respect to the purpose of the experiment. All participants gave informed consent prior to being included in the study.

Stimuli

For the normal face condition, we used average male and female faces (with a neutral facial expression), generated by morphing many different faces (Gruendl, M., http://www.beautycheck.de). Three different types of scrambled male and female faces were created by shifting the positions of each individual part (e.g., nose, mouth, eyes) from their canonical positions (e.g., mouth below the nose) so as to distort the face-specific configuration (see Fig. 5a, b) while preserving the local orientation of face parts (e.g., horizontal mouth). Local orientation of face parts was preserved to render the target stimulus identical in both normal and scrambled face conditions.

The control stimuli were houses. These stimuli were chosen for the purpose of matching faces with comparably complex, socially elevant and familiar items (Fig. 5c). Normal house stimuli were represented by a house with a door and three windows. Scrambled houses were created by changing the positions of the door and windows, while preserving the local orientation of the house parts.

Procedure

The procedure of the Experiment 2 was identical to Experiment 1, except for the use of a more realistic set of stimuli. The task was to discern the presence (vs. absence) of a target stimulus. The target was blue colored mouth amongst green colored lips for the face condition (see Fig. 5a, b), and a blue-colored door amongst green-colored doors for the house condition (Fig. 5c). The order of the conditions was counterbalanced across participants. Search displays always contained six face or house stimuli, and the target (blue lips or blue door) could appear only in one of them (Fig. 6).

Results

One participant made substantially more errors (>10 %) than average error rates (6 %), and was excluded from further analyses. Figure 7 shows the mean RTs and errors for across experimental conditions in Experiment 2. Similar to Experiment 1, target-present RTs were overall faster than target-absent RTs (466 vs. 518, $p < 0.01$). No other effects proved significant (all ps > 0.05), indicating that normal and scrambled face and house stimuli were processed with comparable speed across experimental conditions.

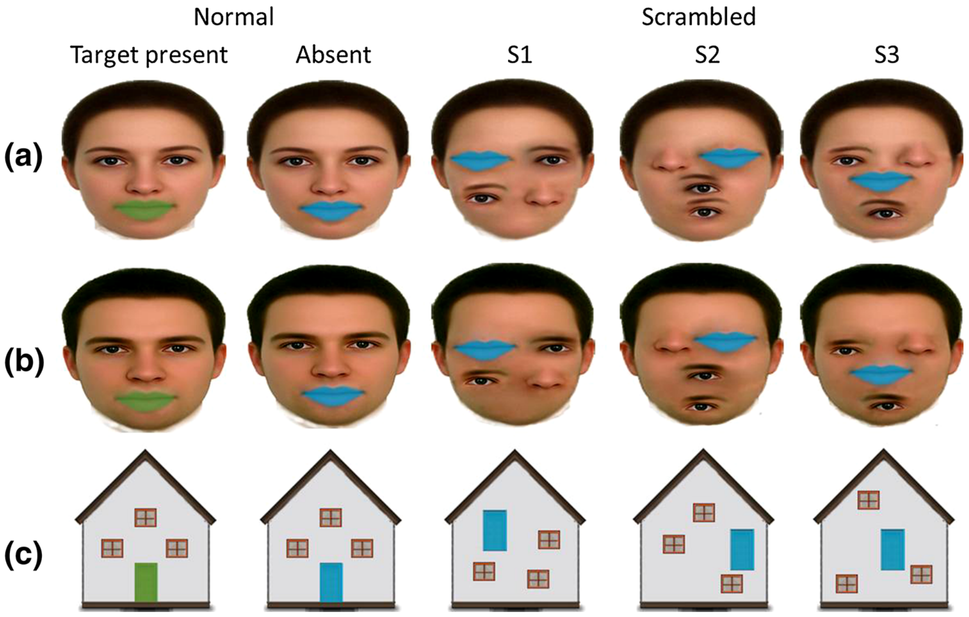


Fig. 6 Illustrates the **a** female and **b** male faces used in Experiment 2, along with **c** composite house stimuli. In the experiment, an array of six equidistantly spaced faces (or houses) was presented in the search display

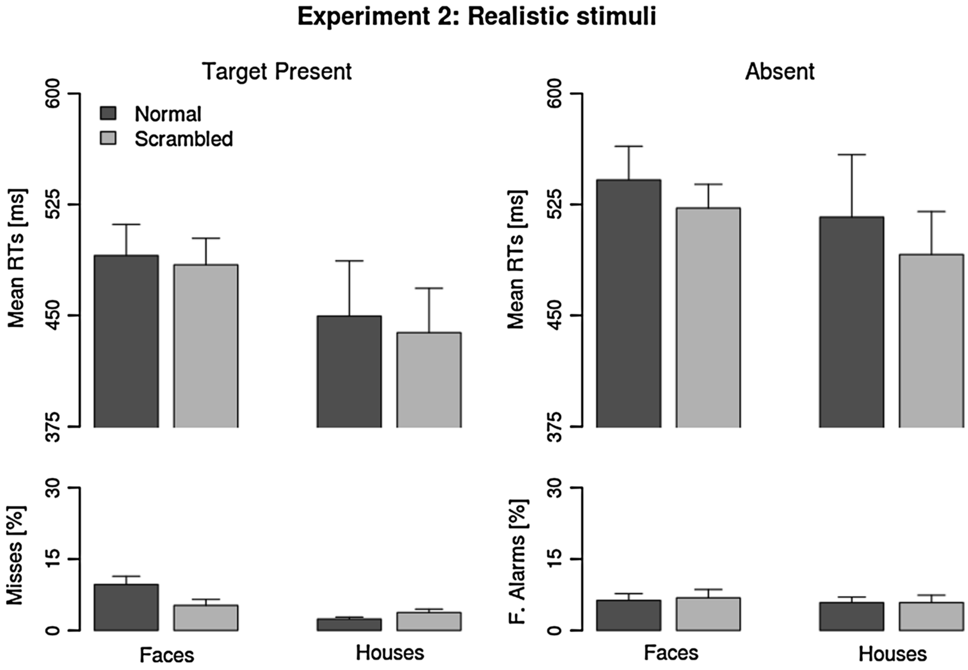


Fig. 7 Reaction times and error rates across experimental conditions of Experiment 2. *Whiskers* denote 1 SEM

In contrast to Experiment 1 in which the rate of misses was smaller than that of false alarms, it was comparable in Experiment 2 (misses vs. false alarms: 5 vs. 7 %, $p = 0.69$) suggesting an unbiased response criterion in Experiment 2. Furthermore, responses were less accurate for faces than for houses (8 vs. 5 %, $p < 0.05$), while the main effect of stimulus configuration failed to reach significance. Only the three-way interaction between target type, stimulus type and configuration proved significantly [$F(1,18) = 5.09$, $\eta_p^2 = 0.22$, $p < 0.05$] attributable to the fact that LFS effect for faces was significant for target-present trials (5 % LFS, $p < 0.05$) only. No stimulus configurations effects (LFS or LFE) proved significant in other conditions (all ps > .05). In summary, in Experiment 2, no effects of configuration were observed in the mean RTs. For the errors, by contrast, there was a LFS effect for faces on target-present trials.

Discussion

Experiment 2 examined whether the LFS and LFE effects observed in Experiment 1 for schematic face and comparable control stimuli (Kanizsa) would also be obtained if participants are presented with more realistic and familiar (face and house) stimuli. The answer is affirmative, with the results replicating the pattern of effects observed in Experiment 1: detecting a local target within a normal face configuration was harder than detecting a target in a scrambled configuration (LFS effect). Importantly, in contrast to Experiment 1, there was no evidence that participants adopted a biased response criterion, and yet the LFS effect for faces was observed—again arguing against post-perceptual processes related to response selection being the origin of the LFS effect.

General discussion

The aim of the present study was to examine what kind of stimulus properties is actually encoded by the hypothesized face detectors. The results strongly suggest that they process only global face properties, that is: the 'faceness' of a stimulus. Specifically, the results revealed processing of task-relevant stimuli to be worse when embedded within a face, relative to a scrambled configuration—that is, there was a LFS effect. Importantly, the LFS effect was specific to face stimuli, and absent—or there was even LFE—for other types of stimuli. These findings are even more striking considering that our paradigm permitted fast and accurate responses simply based on processing local stimulus properties (shape and color in Experiments 1 and 2, respectively), that is, disregarding the global configurations altogether. Since normal and scrambled configurations consisted of identical local elements, adoption of such a local processing mode would predict no configuration effects. However, at variance with this prediction, configuration effects were observed in all our experiments—suggesting that participants were not able to solely use simple feature information and disregard the global configuration for solving the task.

Overall, the present findings are in line with other studies that demonstrated face processing to be holistic in nature (e.g., Leder & Carbon, 2005; Tanaka & Farah, 1993). Leder and Carbon (2005) reported a similar LFS effect, albeit in a different task, by showing that local face elements are recognized poorly when presented in a global context (i.e., whole face) relative to isolated presentation. Our results support these findings and go beyond them by showing that global face perception occurs early during visual processing (i.e., with stimuli presented for only 200 ms), possibly reflecting the existence of specific detectors tuned to faces—but not of detectors for other complex objects (e.g., Kanizsa figures, houses), as these are not vital for successful adaptation and social interaction. An alternative to postulating a face-specific mechanism responsible for the LFS effect observed in the present study might be that LFS effects are specific for a particular level of task difficulty or a particular response bias. However, while these factors, related to response selection processes, might explain the LFS effect for faces in a single experiment, the fact that we consistently observed similar effects across a set of experiments that differed in overall task difficulty and response bias renders the response-based explanation of the LFS effect unlikely. Taken together, the LFS effect for schematic and realistic faces and the LFE effect for Kanizsa and realistic house stimuli would then indicate that faces are indeed processed as a single feature and that face detectors are tuned to global face properties. This assumption is supported by recent findings from behavioral (Goffaux & Rossion, 2006) and electrophysiological studies (Halit, de Haan, Schyns, & Johnson, 2006), which indicate that the face-specific processing is primarily based on the spectrum of low spatial frequencies (i.e., reflecting global stimulus properties), further suggesting that faces are processed differently already at the early, perceptual processing stages.

The proposal of globally tuned face detectors would also explain why previous studies failed to find a pop-out effect for faces when these were presented among homogeneous face distractors. Search displays consisting only of face-like configurations would result in face-detectors signaling the presence of something resembling a face at multiple locations, as a result of which attention would not be guided efficiently to the location of the face target. Finally, one of the major criticisms leveled at findings of pop-out search for global face configurations was that this effect is attributable to local differences between faces and other stimuli. In the light of the present findings, this argument would no longer be valid: because the local features of faces are processed less efficiently, compared to control stimuli, it is unclear how less efficiently processed local face features could lead to pop-out, relative to local features of other objects (e.g., a car) which do not produce pop-out.

Rather than assuming that performance in our task was mediated by face detectors tuned to global stimulus attributes, one could argue that the obtained LFS effect reflects the dynamics of post-selective visual processing. In particular, the LFS effect could arise from processes related to memory for previously selected and discarded locations. That is, having to process faces might influence the memory of previously selected and, upon focal-attentional checking, rejected locations in such a way that the chance of re-selecting a previously attended global configuration is greater for normal relative to scrambled faces. In brief, faces might induce weaker inhibition of return (IOR, Klein, 2000; Posner & Cohen, 1984), which in turn causes repeated search through items that have already been checked, thus slowing RTs in the face condition. In a recent study, Theeuwes and Van der Stigchel (2006) examined this hypothesis. Participants were presented with two photographs—one of them was always a face, the other always a non-face stimulus—located on the left and right sides of the central fixation cross. After a variable interval, a central pointer indicated the direction (i.e., to the right or the left of the cross) to which participants had to make a saccade. The results showed that participants were slower in making a saccade to the position that had previously contained the face stimulus, suggesting that faces actually generate a stronger IOR effect. This pattern is at variance with the alternative account on which weaker IOR for faces is the source of stronger face-specific LFS effects.

On the other hand, the LFS effect might be caused by the fact that it takes longer to disengage one's attention from a selected face, relative to other stimuli (Devue, Belopolsky, & Theeuwes, 2011; Bindemann, Burton, & Jenkins, 2005). Accordingly, search through an array of faces would be slower, relative to arrays of non-face stimuli. However, longer focal-attentional processing of face stimuli should also result in improvements in performance, predicting more accurate (local) target detection for normal face configurations—relative to scrambled configurations, which would be analyzed more briefly. Our results do not support this explanation, since we observed both slower and less accurate performance for face-like targets relative to scrambled targets. Thus, all arguments considered, the LFS effect observed for faces but not for non-face stimuli (i.e., the fact, that local features of faces were processed less effectively compared to those of non-faces) in the present study is likely to originate from the early processes of stimulus encoding and detection—consistent with the face detector hypothesis.

Conclusion

The present study revealed that processing of local visual information is hampered when it is part of a normal, as compared to a scrambled, face. Comparable results were observed with both schematic and realistic faces. This LFS effect for faces adds support to Hershler and Hochstein's (2005, 2006) proposal that faces are encoded by special visual face detectors, that is, as a basic visual feature similar to other features in the color, orientation, motion, etc. dimensions. Finally, these detectors seem to be tuned to the general global face configuration, rather than local facial features.

References

Barragan-Jason, G., Besson, G., Ceccaldi, M., & Barbeau, E. J. (2013). Fast and famous: looking for the fastest speed at which a face can be recognized. *Frontiers in Psychology, 4*, 100. doi:10.3389/fpsyg.2013.00100.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development, 23*, 223–229. doi:10.1016/S0163-6383(01)00037-6.

Bindemann, M., Burton, A. M., Hooge, I. T., Jenkins, R., & de Haan, E. H. (2005a). Faces retain attention. *Psychonomic Bulletin and Review, 12*(6), 1048–1053. doi:10.3758/BF03206442.

Bindemann, M., Burton, A. M., & Jenkins, R. (2005b). Capacity limits for face processing. *Cognition, 98*(2), 177–197. doi:10.1016/j.cognition.2004.11.004.

Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: do faces pop out? *Perception, 26*(12), 1555–1570. doi:10.1068/p261555.

Carbon, C. C. (2011). The first 100 milliseconds of a face: on the microgenesis of early face processing. *Perceptual and Motor Skills, 113*(3), 859–874. doi:10.2466/07.17.22.PMS.113.6.859-874.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22*(2), 225–271. doi:10.1016/0010-0285(90)90017-X.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just. *Journal of Vision, 10*(4), 16.1–17. doi:10.1167/10.4.16.

Devue, C., Belopolsky, A. V., & Theeuwes, J. (2011). The role of fear and expectancies in capture of covert attention by spiders. *Emotion, 11*(4), 768–775. doi:10.1037/a0023418.

Foster, D. H., & Ward, P. A. (1991). Horizontal-vertical filters in early vision predict anomalous line-orientation identification frequencies. *Proceedings of the Royal Society B: Biological Sciences, 243*(1306), 83–86. doi:10.1098/rspb.1991.0014.

Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—holistic face perception is supported by low spatial frequencies. *The Journal of Experimental Psychology: Human Perception and Performance, 32*(4), 1023–1039. doi:10.1037/0096-1523.32.4.1023.

Halit, H., de Haan, M., Schyns, P. G., & Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research, 1117*(1), 154–161. doi:10.1016/j.brainres.2006.07.059.

Hershler, O., & Hochstein, S. (2005). At first sight: a high-level pop out effect for faces. *Vision Research, 45*(13), 1707–1724. doi:10.1016/j.visres.2004.12.021.

Hershler, O., & Hochstein, S. (2006). With a careful look: still no low-level confound to face pop-out. *Vision Research, 46*(18), 3028–3035. doi:10.1016/j.visres.2006.03.023.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in Dynamic Scenes. *Visual Cognition, 12*(6), 1093–1123. doi:10.1080/13506280444000661.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual-attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 1254–1259. doi:10.1109/34.730558.

Jolicoeur, K. (1994). Impact of quality of the image, orientation, and similarity of the stimuli on visual search for faces. *Perception, 23*(1), 95–122. doi:10.1068/p230095.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition, 68*, 1–11. doi:10.1016/S0010-0277(98)00035-3.

Kietzmann, T. C., & König, P. (2010). Perceptual learning of parametric face categories leads to the integration of high-level class-based information but not to high-level pop-out. *Journal of Vision, 10*(13), 1–14. doi:10.1167/10.13.20.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences, 4*(4), 138–147. (PII: S1364-6613(00)01452-2).

Leder, H., & Carbon, C. C. (2005). When context hinders! Learn-test compatibility in face recognition. *The Quarterly Journal of Experimental Psychology Section A. Human Experimental Psychology, 58*(2), 235–250. doi:10.1080/02724980343000936.

Lobmaier, J. S., Bolte, J., Mast, F. W., & Dobel, C. (2010). Configural and featural processing in humans with congenital prosopagnosia. *Advances in Cognitive Psychology, 6*, 23–34. doi:10.2478/v10053-008-0074-4.

Murray, J. (2002). Evidence from visual search for holistic processing of inverted faces. *JoV, 2*(7), 65. doi:10.1167/2.7.605.

Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception, 22*(11), 1287–1298. doi:10.1068/p221287.

Peirce, J.W. (2007). PsychoPy - psychophysics software in Python. *Journal of Neuroscience Methods, 162*, 8-1. doi:10.1016/j.jneumeth.2006.11.017.

Peirce, J.W. (2009) Generating stimuli for neuroscience using PsychoPy. Front. *Neuroinform*, *2*, 10. doi:10.3389/neuro.11.010.2008.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bouwhuis (Eds.), *Chapter in attention and performance X* (pp. 531–556). USA: Erlbaum.

Powell, N., & Humphreys, B. (1984). Proportions of the aesthetic face. New York, NY: Thieme-Stratton.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research, 51*, 333–342. doi:10.1016/j.visres.2010.11.014.

Rouw, R., & Gelder, B. D. (2002). Impaired face recognition does not preclude intact whole face perception. *Visual Cognition, 9*(6), 689–718. doi:10.1080/13506280143000223.

Schwaninger, A. L., Lobmaier, J. S., & Collishaw, S. M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition. *Lecture Notes in Computer Science, 2525*, 643–650. doi:10.1007/3-540-36181-2_64.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology A, 46*(2), 225–245. doi:10.1080/14640749308401045.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, *25*, 583-592. doi:10.3758/BF03211301.

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: evidence from Inhibition-of-return. *Visual Cognition, 13*, 657–665. doi:10.1080/13506280500410949.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review, 95*(1), 15–48. doi:10.1037/0033-295X.95.1.15.

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471–491. doi:10.1111/j.2044-8295.1988.tb02747.x.

VanRullen, R. (2006). On second glance: still no high-level pop-out effect for faces. *Vision Research, 46*(18), 3017–3027. doi:10.1016/j.visres.2005.07.009.

Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598. doi:10.1111/j.1467-9280.2006.01750.x.

Wolfe, J.M., Cave, K. R., Franzel, S.L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology*: Human Perception and Performance, *15*(3), 419-433. doi:10.1037/0096-1523.15.3.419.

Wolfe, J. M., O'Neill, P., & Bennett, S. C. (1998). Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Perception and Psychophysics, 60*(1), 140–156. doi:10.3758/BF03211924.

Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141–145. doi:10.1037/h0027474.