BRIEF REPORT

Face detection differs from categorization: Evidence from visual search in natural scenes

Markus Bindemann · Michael B. Lewis

Abstract In this study, we examined whether the detection of frontal, ¾, and profile face views differs from their categorization as faces. In Experiment 1, we compared three tasks that required observers to determine the presence or absence of a face, but varied in the extents to which participants had to search for the faces in simple displays and in small or large scenes to make this decision. Performance was equivalent for all of the face views in simple displays and small scenes, but it was notably slower for profile views when this required the search for faces in extended scene displays. This search effect was confirmed in Experiment 2, in which we compared observers' eye movements with their response times to faces in visual scenes. These results demonstrate that the categorization of faces at fixation is dissociable from the detection of faces in space. Consequently, we suggest that face detection should be studied with extended visual displays, such as natural scenes.

The failure to detect faces can lead to the loss of important social information, such as the identity of people in our surroundings or their emotional and attentive state. Our cognitive system seems to guard specifically against such losses, as faces are detected ultra-rapidly (Crouzet, Kirchner, & Thorpe, 2010; Crouzet & Thorpe, 2011) and provide a powerful draw for observers' attention (Langton, Law, Burton, & Scweinberger, 2008; Theeuwes & Van der Stigchel, 2006). These feats emphasize the important status of face detection in social cognition. Despite this, however, this process has received limited attention from psychologists. A convincing theory of human face detection, for example, still does not

M. Bindemann (✉)
School of Psychology, University of Kent,
Canterbury CT2 7NP, UK
e-mail: M.Bindemann@kent.ac.uk

M. B. Lewis
School of Psychology, Cardiff University, Cardiff, UK

exist, and this topic is rarely considered in grand reviews of face perception (see, e.g., Bruce & Young, 2012; Calder, Rhodes, Johnson, & Haxby, 2011; Hole & Bourne, 2010).

A possible reason for this oversight is variation in the general methods with which face detection has been studied, so the specificity of this process is still not inherently clear. Some research has employed paradigms in which faces are cropped from extraneous background and presented in the center of visual displays (e.g., Liu, Harris, & Kanwisher, 2002; Nestor, Vettel, & Tarr, 2008). Other studies have used more complex displays, such as visual search arrays (e.g., Hershler & Hochstein, 2005) and natural scenes (e.g., Bindemann & Burton, 2009).

An empirical question that remains to be resolved is whether this pragmatic divide has theoretical implications for our understanding of face detection. In experimental displays that consist of a single centrally presented stimulus, the process of detection effectively becomes a categorization task, in which observers have to determine whether a fixated stimulus is a face or a nonface object. By contrast, extended visual arrays require observers to determine the presence or absence of a face anywhere in a display. Both types of paradigms therefore rely on face-versus-nonface decisions, but one approach requires the *search* for face stimuli, whereas the other does not. Models of object recognition suggest that detection and categorization are distinct processes (e.g., Driver & Bayliss, 1996; Nakayama, He, & Shimojo, 1995). However, this view has also been contested (Grill-Spector & Kanwisher, 2005), and it is unknown whether such a distinction applies to face processing.

One source of evidence to suggest that this distinction might be appropriate comes from tasks in which observers are required to detect the presence of faces at different locations in the visual field (Hershler, Golan, Bentin, & Hochstein, 2010). Under these conditions, face detection performance is comparable to that for nonface targets when these stimuli are presented close to fixation. However, a detection advantage for faces emerges when the targets appear at greater eccentricities, and particularly when they

are surrounded by further stimuli. These findings therefore provide some initial evidence that different processes are involved in the categorization of centrally presented faces and the detection of these stimuli across the visual field.

We have also obtained indirect evidence for such a distinction in our own work. It has already been shown, for example, that familiar-face recognition (Troje & Kersten, 1999), gender classification (Bindemann, Scheepers, & Burton, 2009), and emotion recognition (Matsumoto & Hwang, 2011) are all unaffected by differences in view (frontal, ¾, and profile) when faces can be located easily onscreen. By contrast, performance declines in profile view when observers have to search for faces in visual scenes (Burton & Bindemann, 2009), which hints, again, that detection might be a process that is separable from other face tasks.

The present study addresses this issue directly, with two contrasting paradigms designed to dissociate these possibilities. We conducted two experiments in which observers were asked to look for faces in different views. In Experiment 1, we used this view manipulation to determine whether the categorization of faces differs from the search for faces in visual scenes. For this purpose, observers were presented with three separate tasks. The first was a categorization task between faces and objects presented at fixation. We compared performance in this paradigm with a detection task, in which observers had to search for faces in visual scenes. We also provided a further task to bridge the gap between these paradigms. In this task, observers were presented with small outtakes from the visual scenes, which were shown in the screen center and might or might not contain a face.

Each of these tasks therefore required observers to determine the presence or absence of a face, but differed in the extents to which observers had to search for faces to make this decision. The aim here was to determine whether an effect of face view was present during the categorization of stimuli at fixation or only when observers had to search for faces in the first place.

Experiment 1

Method

Participants

A group of 27 students, with a mean age of 20.0 years ($SD = 1.7$), participated in the experiment for course credit.

Stimuli

This experiment consisted of three separate tasks. In Task 1, we assessed face categorization in simple visual displays, whereas in Task 3, we measured face detection in natural scenes. Task 2 was designed to bridge the gap between these paradigms, and therefore is described last.

Task 1: Face categorization in basic displays The stimuli for Task 1 consisted of the faces of 20 models (ten male) and 120 nonface objects. Each person was depicted in five different poses (frontal, ¾ profile left/right, and profile left/right), giving a total of 100 different images. The nonface stimuli consisted of a wide range of objects, including household items, furniture, and food. All images were cropped to remove extraneous background and scaled to a height and width of 180 pixels presented at a resolution of 72 pixels/in. (for examples, see Fig. 1).

Task 3: Face detection in natural scenes The stimuli for Task 3 consisted of photographs of 120 indoor scenes, which were taken from inside houses, apartments, and office buildings. These scenes measured 1,000 (W) × 750 (H) pixels, presented at a screen resolution of 72 pixels/in. For each scene, six versions existed that were identical in all respects, except for the following differences. Five of the versions contained a face, whereas one did not. In face-present scenes, the faces were depicted in a frontal, ¾ profile left/right, or profile left/right view. Applying these manipulations to each of the scenes resulted in a total of 720 different displays, comprising 120 face-absent scenes and 600 face-present scenes.

Crucially, the faces in these scenes were the same photographs that were used in Task 1 (see Fig. 1c). The location of these photographs was unpredictable across the scenes. In addition, the faces varied in size across the scenes, ranging from 0.08 % to 1.73 % of the total scene area, to ensure that participants could not adopt a simple search strategy based on target size (for further details, see Burton & Bindemann, 2009).

Task 2: Face categorization in small scenes This task was designed to bridge the gap between the face/nonface categorization paradigm (Task 1) and the search for faces in full scenes (Task 3). The stimuli therefore consisted of the scene areas that contained the target faces (see Fig. 1b). These scene segments measured 180 pixels in height and width and were presented in the center of a plain white background. The scale of these stimuli was preserved, so that the faces were presented at the same size as in the full scenes.

To provide a set of nonface stimuli to make up the task demands, the corresponding areas from the face-absent scenes were used. In total, therefore, 720 different displays were presented, comprising 120 face-absent and 600 face-present images, in which a face was shown in a frontal view (120 images), a ¾ left or right view (120 images each), or a profile left or right view (120 images each).
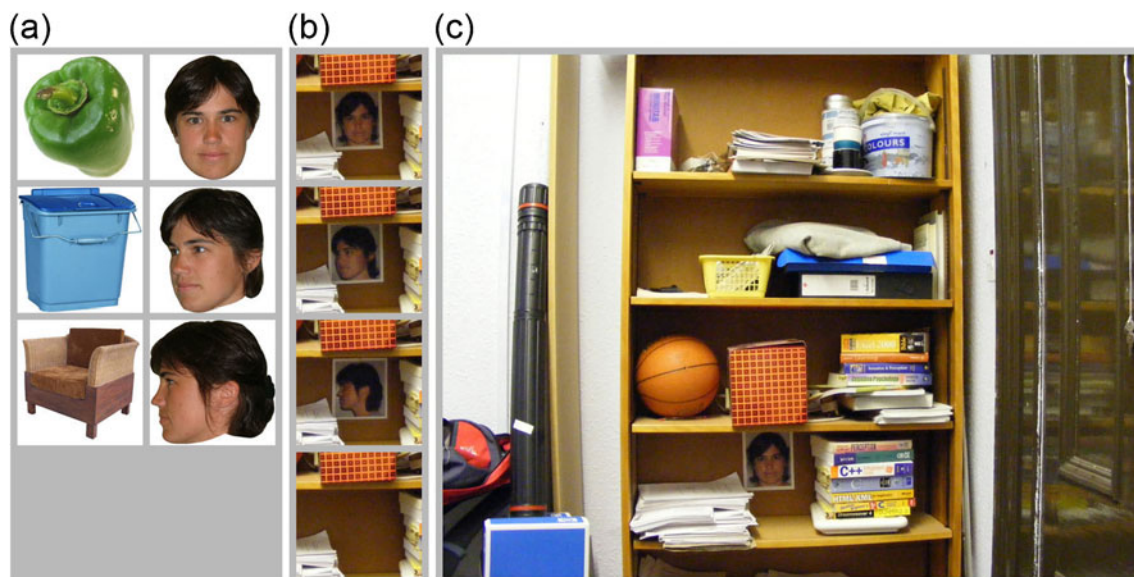
Fig. 1 Examples of the face and nonface stimuli for Task 1 (**a**) and Task 2 (**b**), as well as an example of a face-present scene for Task 3 (**c**), in Experiment 1

Procedure

In each of the tasks, participants were shown 360 randomly intermixed trials, consisting of 240 nonface and 120 face trials. The face trials consisted of 40 stimuli for each of the three view conditions (frontal, ¾ view, profile view). For the ¾ and profile views, these comprised 20 left- and 20 right-facing stimuli. In the full- and partial-scene tasks, the different face views were rotated around the scene images, so that each face-present scene was only shown once to each participant. Overall, however, the presentation of the scenes was counterbalanced across participants, so that each face view appeared in each scene an equal number of times.

In each of the tasks, a trial began with a fixation cross for 1 s, followed by a stimulus, which was displayed until a response was registered. Participants were briefed about the different face views prior to the experiment and were asked to make speeded responses concerning whether or not a face was present. The running order of the three tasks was counterbalanced across participants.

Results

In all three tasks, the average correct response times were calculated for the three face views. These results are displayed in Table 1 and show that response times were generally comparable for Tasks 1 (face categorization in basic displays) and 2 (face categorization in small scenes) but were slower in Task 3 (face detection in natural scenes). In addition, performance also appeared to be matched evenly across face views in Tasks 1 and 2, but response times were notably slower for profile faces than for frontal and ¾ views in Task 3.

These observations were confirmed by a 3 × 3 within-subjects analysis of variance (ANOVA), which showed a main effect of task, $F(2, 52) = 94.46$, $p < .001$, $\eta_p^2 = .78$. Tukey HSD comparisons showed that this effect arose from slower response times on Task 3 than on Tasks 1 and 2, both qs ≥ 16.23, ps ≤ .001, ds ≥ 1.70, whereas Tasks 1 and 2 did not differ, $q = 1.15$. In addition, the ANOVA also showed an effect of face view, $F(2, 52) = 3.58$, $p < .05$, $\eta_p^2 = .12$, and an interaction between both factors, $F(2, 52) = 4.44$, $p < .01$, $\eta_p^2 = .15$. An analysis of simple main effects showed a difference in response times across face views in Task 3, $F(2, 52) = 11.28$, $p < .001$, $\eta_p^2 = .30$, but not in Task 1 or 2, both Fs < 1. A comparison of the face views in Task 3 showed that response times were evenly matched for frontal and ¾ views, $q = 1.32$, but were slower for profile faces than for frontal and ¾ views, both qs ≥ 5.05, ps ≤ .01, ds ≥ 0.33.

A 3 × 3 ANOVA of the accuracy of responses also showed a main effect of task, $F(2, 52) = 20.50$, $p < .001$, $\eta_p^2 = .44$, which reflected lower accuracy in Task 3 than in Tasks 1 and 2, both qs ≥ 6.64, ps ≤ .001, ds ≥ 0.78, whereas Tasks 1 and 2 did not differ from each other, $q = 2.01$. The ANOVA also revealed a main effect of face view, $F(2, 52) = 12.52$, $p < .001$, $\eta_p^2 = .33$, reflecting lower accuracy in the classification of profile faces than for frontal and ¾ views, both qs ≥ 4.42, ps ≤ .01, ds ≥ 0.22. By comparison, accuracy for the frontal and ¾ view conditions did not differ, $q = 2.58$. The interaction of task and view was not significant, $F(2, 52) < 1$.

Discussion

This experiment replicated previous research by showing that the detection of profile faces is slower in natural visual

Table 1 A summary of the mean response times and accuracy for the three tasks in Experiment 1

	Face View			No Face
	Frontal	3/4	Profile	
Response Times (in ms)				
Task 1 (Categorization)	533 (137)	514 (129)	504 (111)	538 (151)
Task 2 (Categorization)	539 (115)	545 (123)	555 (108)	554 (121)
Task 3 (Search)	929 (250)	900 (246)	1,043 (428)	1,580 (888)
Accuracy (%)				
Task 1 (Categorization)	94.8 (3.9)	95.9 (5.2)	94.4 (5.0)	98.7 (2.0)
Task 2 (Categorization)	92.4 (10.3)	93.4 (9.3)	90.3 (9.7)	97.4 (6.5)
Task 3 (Search)	82.7 (16.1)	83.8 (15.4)	79.6 (14.6)	96.4 (8.5)

Parentheses show the standard deviations of the means. The data for "no face" responses are reported for completeness but were not included in the analysis.

scenes than is detection of faces in frontal and ¾ views (Burton & Bindemann, 2009). Crucially, this was observed in a context that did not show the same effect of view when faces were presented centrally, so that the search component of the detection task was eliminated. Moreover, this was observed with faces that were shown on a plain background (Task 1) and in small scenes (Task 2). In the latter condition, the faces were an exact match of those shown in the full scenes, in terms of their size, color, and contrast, as well as in the immediate surrounding visual context. The difference between these tasks therefore shows that the categorization of a centrally presented stimulus as a face and the search for faces in extended visual displays can yield rather different outcomes. This indicates that categorization and detection are dissociable face processes.

Experiment 2

Although the effect of view was only observed in Task 3 of Experiment 1, the notion that this effect does, in fact, occur during the *search* process was only inferred from the response times on the basis that a similar pattern was not found for centrally presented faces in Tasks 1 and 2. In Experiment 2, we investigated this issue more directly by recording observers' eye movements during the search for faces in natural scenes. Eye movements are closely associated with the allocation of attention around stimulus displays (Deubel & Schneider, 1996) and provide a sensitive online measure for visual search tasks (Liversedge & Findlay, 2000), including the detection of faces (Crouzet et al., 2010). We therefore contrasted the effect of view on the time required to first fixate a face in a visual scene, which provides an immediate eyetracking measure of the duration of the search process, with the effect on the time to respond to a face. By comparing these measures, it was possible to determine more directly whether the effect of view arises during visual search for possible face candidates or at a subsequent, categorization stage that might take place to confirm that a looked-at stimulus really is a face.

Method

Participants

A group of 18 students, with a mean age of 21.7 years ($SD =$ 3.5), participated in the experiment for course credit.

Stimuli and procedure

The stimuli and procedure were the same as in Task 3 of Experiment 1, except for the following changes. Participants' eye movements were recorded during the detection task. The stimuli were therefore displayed on a 21-in. color monitor that was connected to an EyeLink 1000 desk-mounted eyetracking system running at a 500-Hz sampling rate. The scenes were presented at a size of 1,000 (W) × 750 (H) pixels at a screen resolution of 66 pixels/in. Viewing was binocular, but only the participants' left eye was tracked. To calibrate the tracker, the standard nine-point EyeLink calibration procedure was used initially and repeated every 60 trials.

In the experiment, each trial began with a central fixation dot, which was used to perform an automatic drift correction. A stimulus was then presented until a response was registered. Participants made speeded keypress responses concerning whether or not a face was present. Each participant was given 360 trials, comprising 120 face-present (40 each for frontal, ¾, and profile view targets) and 240 face-absent trials, in a randomly intermixed order.

Results and discussion

The mean response times and accuracy for all views are shown in Table 2. A one-factor ANOVA of the response time data showed an effect of face view, $F(2, 34) = 9.38$, $p <$ .001, $\eta_p^2 = .36$, reflecting longer response times for profile faces than for frontal and ¾ views, both $qs \geq 3.84$, $ps \leq .05$, $ds \geq 0.47$. In contrast, the difference between frontal and ¾ views was not significant, $q = 2.22$. A corresponding pattern was found for accuracy. A one-factor ANOVA showed an

Table 2 A summary of the response and eye movement measures in Experiment 2

	Face View			No Face
	Frontal	3/4	Profile	
Response time	847 (103)	881 (132)	940 (118)	1,680 (425)
Accuracy	92.5 (4.0)	92.4 (6.7)	88.8 (6.1)	99.1 (0.8)
Time to fixation	593 (108)	588 (114)	704 (188)	—
Number of fixations	2.4 (0.5)	2.4 (0.4)	2.8 (0.5)	—

Parentheses show the standard deviations of the means. The data for "no face" responses are reported for completeness but were not included in the analysis.

effect of face view, $F(2, 34) = 6.15$, $p < .01$, $\eta_p^2 = .27$, due to lower accuracy in the profile face condition than for the frontal and ¾ views, both qs ≥ 4.21, ps ≤ .05, ds ≥ 0.56, whereas the difference between frontal and ¾ views was not significant, $q = 0.16$.

In addition, two measures were extracted from the eye movements, corresponding to the average time that was required to first fixate a face in a visual scene (*time to fixation*) and the average number of fixations that were required to do so (*number of fixations*). Both measures provide a direct index of the search effort that is required to locate a face and are summarized in Table 2.

A one-factor ANOVA of the time-to-fixation data showed an effect of face view, $F(2, 34) = 11.67$, $p < .001$, $\eta_p^2 = .41$, due to longer search times for profile faces than for frontal and ¾ views, both qs ≥ 5.77, ps ≤ .001, ds ≥ 0.72, whereas the difference between the frontal and ¾ views was not significant, $q = 0.27$. A similar pattern was found for number of fixations, $F(2, 34) = 13.02$, $p < .001$, $\eta_p^2 = .43$, with observers requiring more fixations to locate profile faces, relative to faces in frontal and ¾ views, both qs ≥ 6.01, ps ≤ .001, ds ≥ 0.85, whereas the difference between the frontal and ¾ views was not significant, $q = 0.44$.

This experiment therefore provides more direct evidence that the effect of face view arises during the *search* for faces in visual scenes. As in Experiment 1, observers were slower to locate profile than frontal and ¾ views of faces. Crucially, however, this same effect was found in the time that observers required to first fixate a face in visual scenes. Experiment 2 therefore indicates that the view effect is genuinely a *search* effect, which arises while observers try to locate faces in visual scenes.

General discussion

In this study, we examined whether the search for faces in natural scenes differs from the categorization of visual stimuli as face and nonface objects. In Experiment 1, an effect of face view was found when observers searched for faces in extended scenes, but not when small segments of scenes or isolated images of faces and nonface objects were presented in a central location. Note that these differences are unlikely

to reflect general task difficulty: Although the full-scene stimuli yielded the longest response times in Experiment 1, other categorization tasks, such as familiar-face recognition, yield similarly long response times but do not show view effects (Troje & Kersten, 1999). Finally, to analyze these findings further, observers' eye movements were recorded during the full-scene task in a second experiment. This confirmed that the view effect arises during the search for faces.

Overall, our results therefore indicate that the process of detection, whereby faces have to be searched for in visual scenes, differs from the categorization of face and nonface images. These findings converge with previous experiments that have already hinted at this distinction (Hershler et al., 2010), but they provide more direct evidence by contrasting categorization with visual search in natural scenes and by assessing observers' search behavior with eye movements. These results are of theoretical importance for showing that these processes are dissociable. Moreover, our data suggest that if we wish to understand face detection per se, this process should be studied with extended visual displays, such as natural scenes.

At present, no theories of human cognition provide an adequate account of how faces are detected in our visual field (see, e.g., Hershler & Hochstein, 2005; Lewis & Edmonds, 2005; Lewis & Ellis, 2003), despite the fact that this is the entry point for all other tasks with faces. Even the cause of the effect of face view, which was used to dissociate face detection from categorization here, has so far resisted explanation (Burton & Bindemann, 2009). Initially, it seemed plausible that the eye regions play an important role in face detection (Lewis & Edmonds, 2003) and contribute to the view effect (Burton & Bindemann, 2009). However, this explanation is perhaps less likely now, considering that eye gaze cannot be perceived outside of focal attention (Burton, Bindemann, Langton, Schweinberger, & Jenkins, 2009). The speed of face detection suggests that this process might instead rely on a "quick and dirty" processing strategy that relies on more salient visual cues to locate possible face candidates (Crouzet et al., 2010).

One possibility for such a strategy could be based on a face-shaped color template. This notion is born out of the observation that color information can be processed very rapidly (e.g., Treisman, 1993) and skin-color tones can be an

effective first step for detecting likely face candidates (Bindemann & Burton, 2009). This notion receives further support from the finding that faces can be detected rapidly when internal or external visual features are selectively removed, as long as an oval, face-shaped color template is retained, but not when this template is disrupted by image scrambling (Hershler & Hochstein, 2005). These findings might explain the effect of view in face detection, which arises primarily from information carried in the top half of a face (Burton & Bindemann, 2009). In profile faces, much of this area is occluded by hair (see Fig. 1). If this disrupts the diagnostic oval shape of faces, this could perhaps produce the detection disadvantage that is found for profile views in natural scenes.

This account is clearly speculative, and other theories of face detection have been proposed elsewhere (e.g., Hershler & Hochstein, 2005; Lewis & Edmonds, 2005). These theories do not explicitly deal with variation in the detectability of different face views, but are united in trying to explain this visual process. The main aim of this research is to demonstrate that face detection is indeed an important process in its own right, and is dissociable from the categorization paradigms that are frequently used in face research. We hope that this finding provides an imperative for further progress in this field.

References

Bindemann, M., & Burton, A. M. (2009). The role of colour in human face detection. *Cognitive Science, 33*, 1144–1156.

Bindemann, M., Scheepers, C., & Burton, A. M. (2009). Viewpoint and center of gravity affect eye movements to human faces. *Journal of Vision, 9*(2), 1–16.

Bruce, V., & Young, A. W. (2012). *Face perception*. London, UK: Psychology Press.

Burton, A. M., & Bindemann, M. (2009). The role of view in human face detection. *Vision Research, 49*, 2026–2036.

Burton, A. M., Bindemann, M., Langton, S. R. H., Schweinberger, S. R., & Jenkins, R. (2009). Gaze perception requires focused attention: Evidence from an interference task. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 108–118.

Calder, A. W., Rhodes, G., Johnson, M. H., & Haxby, J. V. (2011). *The Oxford handbook of face perception*. New York, NY: Oxford University Press.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades towards faces: Face detection in just 100 ms. *Journal of Vision, 10*(4), 1–17.

Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers in Psychology, 2*(342), 1–9.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36*, 1827–1837.

Driver, J., & Baylis, G. C. (1996). Edge-assignment and figure–ground segmentation in short-term visual matching. *Cognitive Psychology, 31*, 248–306.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16*, 152–160.

Hershler, O., Golan, T., Bentin, S., & Hochstein, S. (2010). The wide window of face detection. *Journal of Vision, 10*(10), 1–14.

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop-out effect for faces. *Vision Research, 45*, 1707–1724.

Hole, G. J., & Bourne, V. (2010). *Face processing: Psychological, neuropsychological, and applied perspectives*. New York, NY: Oxford University Press.

Langton, S. R. H., Law, A. S., Burton, A. M., & Scweinberger, S. R. (2008). Attention capture by faces. *Cognition, 107*, 330–342.

Lewis, M. B., & Edmonds, A. J. (2003). Face detection: Mapping human performance. *Perception, 32*, 903–920.

Lewis, M. B., & Edmonds, A. J. (2005). Searching for faces in scrambled scenes. *Visual Cognition, 12*, 1309–1336.

Lewis, M. B., & Ellis, H. D. (2003). How we detect a face: A survey of psychological evidence. *International Journal of Imaging Systems and Technology, 13*, 3–7.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience, 5*, 910–916.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4*, 6–14.

Matsumoto, D., & Hwang, H. S. (2011). Judgements of facial expressions of emotion in profile. *Emotion, 11*, 1223–1229.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn & D. N. Oshern (Eds.), *An invitation to cognitive science: Visual cognition* (pp. 1–70). Cambridge, MA: MIT Press.

Nestor, A., Vettel, J. M., & Tarr, M. J. (2008). Task-specific codes for face recognition: How they shape the neural representation of features for detection and individuation. *PLoS One, 3*, e3978. doi:10.1371/journal.pone.0003978

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition, 13*, 657–665.

Treisman, A. (1993). The perception of features and objects. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Awareness, selection, and control* (pp. 5–35). Oxford, UK: Oxford University Press.

Troje, N. F., & Kersten, D. (1999). Viewpoint-dependent recognition of familiar faces. *Perception, 28*, 483–487.