



Kent Academic Repository

Pongakkasira, Kaewmart and Bindemann, Markus (2015) *The shape of the face template: Geometric distortions of faces and their detection in natural scenes*. *Vision Research*, 109 (A). pp. 99-106. ISSN 0042-6989.

Downloaded from

<https://kar.kent.ac.uk/53146/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1016/j.visres.2015.02.008>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

The shape of the face template:
Geometric distortions of faces and their detection in natural scenes

Kaewmart Pongakkasira & Markus Bindemann
School of Psychology, University of Kent, Canterbury, UK

Correspondence to:

Kaewmart Pongakkasira, School of Psychology, University of Kent, CT2 7NP, UK

Email: kp281@kent.ac.uk

Tel: 0044 (0) 1227 823087

Fax: 0044 (0) 1227 827030

Word count (main text and references but excluding abstract and figure captions): 4,845

Abstract

Human face detection might be driven by skin-coloured face-shaped templates. To explore this idea, this study compared the detection of faces for which the natural height-to-width ratios were preserved with distorted faces that were stretched vertically or horizontally. The impact of stretching on detection performance was not obvious when faces were equated to their unstretched counterparts in terms of their height or width dimension (Experiment 1). However, stretching impaired detection when the original and distorted faces were matched for their surface area (Experiment 2), and this was found with both vertically and horizontally stretched faces (Experiment 3). This effect was evident in accuracy, response times, and also observers' eye movements to faces. These findings demonstrate that height-to-width ratios are an important component of the cognitive template for face detection. The results also highlight important differences between face detection and face recognition.

Introduction

Human face detection is the process by which observers find faces within the visual environment (see, e.g., Lewis & Edmonds, 2005; Lewis & Ellis, 2003; Tsao & Livingstone, 2008). This process appears to be distinct from subsequent categorization tasks (Bindemann & Lewis, 2013). However, in contrast to other tasks with faces, such as identification (see, e.g., Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990; Burton, Jenkins, Hancock, & White, 2005) and matching (e.g., Burton, White, & McNeill, 2010; Clutterbuck & Johnston, 2002; Johnston & Bindemann, 2013), emotion recognition (e.g., Calder, Burton, Miller, Young, & Akamatsu, 2001; Calder & Young, 2005), or gaze perception (e.g., Bayliss, Pellegrino, & Tipper, 2004; Driver et al., 1999; Jenkins, 2007), face detection has been studied comparatively little in Psychology. This is surprising considering that detection is an important first step for all other tasks with faces.

The available evidence suggests that face detection is automatic (Lewis & Edmonds, 2003, 2005) and very rapid (Crouzet, Kirchner, & Thorpe, 2010; Fletcher-Watson, Findlay, Leekam, & Benson, 2008). This indicates that this process might rely on a “quick and dirty” processing strategy that utilizes salient visual cues to locate likely face candidates (Crouzet & Thorpe, 2011). One possibility for such a strategy could be based on a simple skin-coloured face-shaped template. This idea is based on the finding that skin-colour tones facilitate detection, but only when this is tied to the general shape of a head. Face detection is impaired, for example, when faces are rendered entirely in greyscale or unnatural colours, or when skin-colour tones are preserved in only part of a face (Bindemann & Burton, 2009). Detection performance declines also when the general shape of a face is disrupted by image scrambling (Hershler & Hochstein, 2005). In contrast, face detection appears to be unaffected by some dramatic transformations, such as the removal of the internal facial features (i.e., the

eyes, nose, and mouth), provided that general face-shape and colour information is retained (Hershler & Hochstein, 2005).

Viewed together, these studies suggest that face detection might be underpinned by skin-coloured, face-shaped templates. Beyond these findings, however, the nature of such a template remains largely unexplored. One aspect, for example, that has been preserved in all previous studies in this field is the height-to-width ratio of faces. Considering the impoverished nature of facial stimuli that allow for detection to proceed unhindered (e.g., Bindemann & Burton, 2009; Hershler & Hochstein, 2005), such natural aspect ratios might be particularly important for detection. However, while this idea seems plausible, an interesting discrepancy exists that might also undermine this notion. In tasks that require the *identification* of faces, substantial geometric distortions, which dramatically disrupt the typical height-to-width aspect ratios of faces, do not appear to affect performance. For example, even when faces are stretched vertically to 150% (Bindemann, Burton, Leuthold, & Schweinberger, 2008) or 200% (Hole, George, Eaves, & Rasek, 2002) of their actual size, while the original horizontal dimensions are maintained, the speed and accuracy of recognition is unaffected. This suggests also that face perception can be remarkably insensitive to manipulations that grossly distort stimulus shape.

In this study, we therefore wish to explore how face detection is affected by such geometric distortions, to further investigate the nature of the template that might be used for this process. For this purpose, observers were asked to locate faces in images of natural scenes in a paradigm that is adopted from previous studies (Bindemann & Burton, 2009; Burton & Bindemann, 2009; Bindemann & Lewis, 2013). In contrast to these studies, faces were either presented with their original aspect ratios intact or these ratios were manipulated. The aim here was to examine whether this would affect the efficacy with which faces can be detected, by recording observers' eye movements and response times to faces. If so, this

would suggest that these aspect ratios are an important dimension of a face detection template.

Experiment 1

The first experiment examined how vertical stimulus distortions affect face detection. In this experiment, observers searched natural visual scenes for frontal views of faces, which were either presented in their original aspect ratio or were stretched vertically to increase the height-to-width ratio. Two different stretch conditions were used. In these, either the original height of the face stimuli was preserved but the width was compressed by half, or the original face width was preserved but height was increased to double. These two conditions therefore provide identical height to width ratios (of 2:1), but one is comparable to the original face stimuli by retaining their height, whereas the other retains their width. If detection operates on a face-template that is sensitive to the height-width ratio of faces, then such geometric distortions should impair detection. As a result, observers should be slower to fixate these stretched faces in visual scenes and to make appropriate detection responses.

Method

Participants

Twenty-seven undergraduate students (8 male, 19 female) from the University of Kent, with mean age of 19.7 years ($SD = 2.2$), participated in this experiment for course credit. All reported normal or corrected-to-normal vision.

Stimuli

The stimuli were adopted from previous detection studies (Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Burton & Bindemann, 2009) and consisted of 24-bit RGB

photographs of 120 indoor scenes, which were taken inside houses, apartments and office buildings. These scene images measured 1000 (H) x 750 (W) pixels at a resolution of 72 pixels/inch (subtending a visual angle of $30.5^\circ \times 23.8^\circ$ at a viewing distance of 60 cm). For each scene, four versions were prepared which were identical in all aspects, except for the following differences. Three of these versions contained a photograph of a frontal face. The faces shown in these scenes were of twenty unfamiliar models (ten male, ten female) of white Caucasian origin. To ensure that the face locations were unpredictable throughout the experiment, the scenes were divided into an invisible 3 x 2 grid of six equally-sized rectangles. Across the stimulus set, the faces were equally likely to appear in any of these regions.

Apart from these commonalities, the three versions of these face-present scenes differed in terms of the aspect ratio of the faces. In the *original* face condition, the height-to-width ratios of all faces were preserved. However, the size of the faces was varied across scenes, ranging from 36 (H) x 27 (W) pixels ($1.2^\circ \times 0.9^\circ$ of VA) for the smallest face photograph to 139 x 115 pixels ($4.7^\circ \times 3.9^\circ$) for the largest face image (mean face image dimensions, 58.7 x 47.2 pixels ($2.0^\circ \times 1.6^\circ$); SD, 19.4 x 16.2 pixels ($0.7^\circ \times 0.5^\circ$)). This was done to ensure that participants could not adopt a simple search strategy based on the size of the faces (see Bindemann & Burton, 2009). The height-to-width ratio of these faces was also calculated. Height was measured as the maximum vertical distance between the facial boundary of the chin and the top of the forehead, whereas width was defined as the maximum horizontal distance between the left and right facial boundary by the ears. Across the stimulus set, the height-to-width ratio ranged from 1.08 to 1.75, with a mean of 1.44 (SD = 0.11). This is consistent with the average height-to-width ratio of this ethnic group (Farkas et al., 2005).

In the other two versions of the face-present scenes, these faces were either stretched vertically to twice the original height (i.e., to be 200%), while the horizontal dimensions were

preserved, in the *vertically stretched* condition, or were compressed horizontally by half (i.e., to 50%) while the vertical dimensions were preserved, in the *horizontally compressed* condition. These two conditions therefore provide equivalent height-to-width ratios, but either only match the height or width of the original face stimuli. These manipulations were applied to each of the 120 scenes, resulting in a total of 360 face-present displays. In addition, a fourth version of each scene image was created in which the faces were absent, yielding 120 face-absent scenes. Example stimuli can be seen in Figure 1.

Procedure

In the experiment, participants' eye movements were tracked using an Eyelink II head-mounted eye-tracking system running at 500 Hz sampling rate and SR-Research ExperimentBuilder software. Viewing was binocular but only the participants' dominant eye was tracked. To calibrate the eye-tracker, the standard 9-point Eyelink procedure was used. Thus, participants fixated a series of nine targets on the display monitor. Calibration was then validated against a second presentation of these targets. If the latter indicated poor measurement accuracy (i.e., a mean deviation of more than 1° of participants' estimated eye position from the target), calibration was repeated.

In the experiment, a trial began with an initial drift correction for which participants were required to focus on a central target. A scene stimulus was then shown until a response was registered. Participants were asked to decide whether a face was present or absent in the scene by pressing one of two possible buttons on a standard computer keyboard. Participants were informed in advance that the faces could appear distorted in these scenes. Regardless of this, participants were requested to respond as quickly and as accurately as possible to the faces.

A total of 360 trials was shown to each participant, which consisted of 240 face-absent trials and 120 face-present trials. For face-present trials, 40 scene stimuli were shown in each of the experimental conditions (original, vertically stretched, horizontally compressed). The scene stimuli were rotated around these conditions across participants, so that each scene was shown only once to an observer in any of the face-present conditions. However, the presentation of the scenes was counterbalanced across participants, so that each scene was equally likely to appear in any of the conditions over the course of the experiment. All trials were presented in a randomly intermixed order.

Results

To assess detection performance, observers' accuracy (%) and response times (median correct RTs) were analysed first. This data is provided in Figure 2 and shows that detection accuracy was comparable in the original and the vertically stretched condition but was reduced for horizontally compressed faces. These observations were confirmed by a one-factor within-subject ANOVA which showed a main effect of face type, $F(2,52) = 100.31$, $p < 0.001$, $\eta_p^2 = 0.79$. Post-hoc comparisons using Tukey HSD test showed that accuracy was reduced for horizontally compressed faces compared to their original and vertically stretched counterparts, both $qs \geq 16.60$, $ps < 0.001$, $ds \geq 4.84$. In contrast, performance for original and vertically stretched faces did not differ, $q = 1.40$, $d = 0.51$.

Observers' response times revealed a similar pattern. A one-factor within-subject ANOVA also revealed a main effect of face type, $F(2,52) = 116.59$, $p < 0.001$, $\eta_p^2 = 0.82$. Tukey HSD test showed that original and vertically stretched faces were detected faster than horizontally compressed faces, both $qs \geq 16.80$, $ps < 0.001$, $ds \geq 3.40$. In addition, response times were faster to vertically stretched than original faces, but this difference was not reliable, $q = 3.35$, $d = 1.32$.

In addition, the median time that was required to first fixate the faces in the visual scenes was also analysed. These *search times* were calculated for correct trials only and provide a more direct index of the search effort that is required to detect a face than button presses (i.e., response times). These eye movements were pre-processed by integrating very short fixations (< 80 ms) with the immediately preceding or following fixation if it lay within one degree of visual angle. The rationale for this was that such short fixations typically result from false saccade planning (see Rayner & Pollatsek, 1989).

As expected, search times were considerably faster than observers' button presses but reveal a similar pattern, whereby face detection appeared to be impaired in the horizontally compressed condition (see Figure 2). Accordingly, a one-factor within-subject ANOVA of this data showed a main effect of face type, $F(2,52) = 50.44$, $p < 0.001$, $\eta_p^2 = 0.66$, due to slower response to horizontally compressed faces than their original and vertically stretched counterparts, both $qs \geq 11.86$, $ps < 0.001$, $ds \geq 2.22$ (Tukey HSD). In contrast, the search times for the original and vertically stretched faces did not differ, $q = 0.84$, $d = 0.30$.

Discussion

This experiment examined whether face detection is affected by the vertical distortion of faces. For this purpose, we compared the detection speed and accuracy of unstretched faces, which were presented in their original dimensions, with faces that were stretched vertically or compressed horizontally. Stretching impaired both the speed and accuracy of face detection. However, this effect was obtained only for faces that were "stretched" by compressing their width. In contrast, when faces were stretched to twice their original height, they were detected as well as their unstretched counterparts.

These results therefore appear to be inconclusive regarding the effect of stretching on face detection. However, a simple explanation might exist for the discrepancy between the

horizontally compressed and the vertically stretched condition. These conditions were designed to be comparable to the original stimuli by retaining either the height (in the horizontally compressed condition) or width (in the vertically stretched condition) of these faces. As a result of this manipulation, however, the faces in the different detection conditions differ in terms of their surface area. In the horizontally compressed condition, for example, this area is reduced to half of the original face stimuli, with a corresponding increase in the vertically stretched condition. Surface area is known to affect face detection, whereby smaller faces are more difficult to detect than large faces (Bindemann & Burton, 2009). This raises the possibility that the effect of face stretching was masked in Experiment 1 by the differences in surface area between conditions. It is conceivable, for example, that the detection of vertically stretched faces was also impaired compared to the unstretched originals, but this effect was offset by the increase in surface area in the former condition. This possibility is explored in Experiment 2.

Experiment 2

In Experiment 1, face detection was impaired for horizontally compressed faces, but not for faces that were stretched vertically. These conditions were matched in terms of their height-to-width ratio but differed in the surface area of the face stimuli. This raises the possibility that the effects of face stretching were offset by differences in area size. To dissociate the effects of surface area and stretching, face detection was assessed with four new conditions in Experiment 2. These comprised two conditions in which the original height-to-weight ratios of faces were retained. However, in one of these conditions the faces were presented at the same size as in Experiment 1, while, in the other, the size of the faces was increased to double their surface area. The faces were compared with two stretched conditions. Both of these provided altered height-to-width by stretching faces vertically by

100% relative to the horizontal dimension. However, in one condition, the overall size of the stretched faces was adjusted so that the surface area was equated with the original face stimuli, whereas, in the other, surface area was also doubled. In line with previous findings, we expected a detection advantage for the large face conditions (see Bindemann & Burton, 2009). In addition, if stretching exerts an effect that operates independent of size, then face detection should be impaired in the stretched face conditions.

Method

Participants

Twenty-four undergraduate students (1 male, 23 female) from the University of Kent, with a mean age of 20.1 years ($SD = 3.8$), participated for course credits. None of them had participated in Experiment 1 and all reported normal or corrected-to-normal vision.

Stimuli and procedure

The stimuli were identical to Experiment 1, except for the following changes. In this experiment, four face-present scenes were included. These consisted of the original face stimuli (in the *original* condition) and a corresponding set of scenes, in which the height-to-weight aspect ratio was retained but the size of the faces was adjusted to double the surface area (in the *original large* condition). In addition, two stretched versions were created, in which the height-width ratio was increased by stretching faces vertically by 100% relative to the horizontal dimension. However, in one of these conditions, the face dimensions were adjusted further so that the surface area matched that of the original faces (in the *stretched* condition). In the other condition, stimulus size was increased so that surface area was at twice its original size (in the *stretched large* condition). Applying these manipulations to the

120 original face-present scenes resulted in a total of 480 experimental displays. Example stimuli are shown in Figure 3.

As in Experiment 1, each participant was shown 360 trials in a randomly intermixed order, comprising 120 face-present and 240 face-absent scenes. The face-present trials consisted of 30 scenes in each of the four experimental conditions (original, original large, stretched, stretched large). As in Experiment 1, the stimuli were rotated around these conditions across observers, but each scene was equally likely to appear in each condition over the course of the experiment.

Results

The data was analysed as in Experiment 1 and is provided in Figure 4. Accuracy was generally higher in the unstretched than the stretched conditions, and also when the surface area was increased to twice the original size. A 2 (face type: original vs. stretched) x 2 (face area: original vs. large) ANOVA showed a main effect of face type, $F(1,23) = 30.64$, $p < 0.001$, $\eta_p^2 = 0.57$, a main effect of face area, $F(1,23) = 46.12$, $p < 0.001$, $\eta_p^2 = 0.67$, and an interaction between both factors, $F(1,23) = 8.51$, $p < 0.01$, $\eta_p^2 = 0.27$. Analysis of simple main effects revealed an effect of face type for targets with the original area, $F(1,23) = 39.91$, $p < 0.001$, $\eta_p^2 = 0.63$, but not for the two large-area conditions, $F(1,23) = 2.28$, $p = 0.14$, $\eta_p^2 = 0.09$. In addition, a simple main effect of face area was found for original, $F(1,23) = 9.85$, $p < 0.01$, $\eta_p^2 = 0.30$, and stretched faces, $F(1,23) = 41.80$, $p < 0.001$, $\eta_p^2 = 0.65$.

Response times were analysed next. An analogous 2 x 2 ANOVA of this data also showed a main effect of face type, $F(1,23) = 27.03$, $p < 0.001$, $\eta_p^2 = 0.54$, a main effect of face area, $F(1,23) = 128.90$, $p < 0.001$, $\eta_p^2 = 0.85$, and an interaction between factors, $F(1,23) = 5.85$, $p < 0.05$, $\eta_p^2 = 0.20$. Analysis of simple main effects showed an effect of face area for original, $F(1,23) = 33.65$, $p < 0.001$, $\eta_p^2 = 0.59$, and stretched faces, $F(1,23) = 105.29$, $p <$

0.001, $\eta_p^2 = 0.82$. These were complemented by simple main effects of face type for faces in their original size, $F(1,23) = 24.57, p < 0.001, \eta_p^2 = 0.52$, and in a large size, $F(1,23) = 5.74, p < 0.05, \eta_p^2 = 0.20$.

The analysis of eye movements also showed a main effect of face type, $F(1,23) = 15.51, p < 0.001, \eta_p^2 = 0.40$, due to faster search times for unstretched faces, and a main effect of face area, $F(1,23) = 47.51, p < 0.01, \eta_p^2 = 0.67$, with faster search times for the larger faces. The interaction between factors was not significant, $F(1,23) = 0.17, p < 0.68, \eta_p^2 = 0.01$.

Discussion

To provide a stronger test for the notion that face detection is affected by vertical distortions, the surface areas of unstretched and stretched faces were equated in Experiment 2. Moreover, to assess whether the effects of stretching and area are dissociable, we include two area conditions, in which the original surface area of the face stimuli was either preserved or doubled. In line with previous work, a clear effect of face area was found, whereby both unstretched and stretched faces were detected faster in the large area conditions (see Bindemann & Burton, 2009). In addition, a separate effect of stretching was found, whereby faces were detected faster in their original height-to-width ratios than in the stretched conditions. This was evident in response times and eye movements, which indicates that this effect arises during the search for faces.

These findings help to clarify the results of Experiment 1. In that experiment, the stretched faces were equated to their original counterparts either in terms of their height or width. However, this manipulation also resulted in unequal surface areas for the faces across all conditions. As a consequence, it was impossible to separate the effect of face area from stretching. In contrast, Experiment 2 shows clearly that stretching impairs detection

performance when the surface area of faces is controlled across conditions. In contrast to face recognition, which appears to be unaffected by the same geometric distortions (Bindemann, Burton, Leuthold, & Schweinberger, 2008; Hole, George, Eaves, & Rasek, 2002), these results suggest that detection relies on a template that incorporates the typical height-to-width aspect ratios of faces. So far, however, the current experiments have explored this notion only with vertically stretched faces. In a final experiment, we compare vertical and horizontal stretches.

Experiment 3

In contrast to the preceding experiments, which compared faces in their original aspect ratios with vertical stretches, the current experiment included faces that were also stretched horizontally by 100%, to twice of the original face width. Face recognition appears to be unaffected by both types of stretches (Bindemann, Burton, Leuthold, & Schweinberger, 2008; Hole, George, Eaves, & Rasek, 2002). In turn, it is important to assess whether detection is only impaired by vertical or also by horizontal distortions of the typical height-to-width aspect ratios of faces.

Method

Participants

Thirty-two undergraduate students (3 male, 29 female) from the University of Kent, with a mean age of 19.3 years ($SD = 1.0$), participated for course credits. None of these students had participated in the preceding experiments. All reported normal or corrected-to-normal vision.

Stimuli and procedure

The stimuli and procedure were identical to Experiment 2, except for the following changes. In addition to the 120 original face-present scenes, in which faces were presented in their natural height-to-width ratio, two more versions were created of each scene. One of these versions consisted of vertically-stretched faces from Experiment 2, which matched the surface area of the original faces. The other version consisted of horizontally-stretched faces. These faces were prepared in the same manner as their vertically-stretched counterparts, except that the opposite height-to-width ratio was used. This resulted in a total of 360 displays, comprising 120 scenes for each of the face-present conditions (original, vertically stretched, horizontally stretched). Example stimuli are shown in Figure 5.

In the experiment, each observer was shown 240 face-absent and 120 face-present displays (40 displays for each of the original, horizontal stretched, and vertical stretched faces) in a randomly-intermixed order. As in previous experiments, the face stimuli were rotated around the three face-present conditions across observers, so that each face-present scenes was only encountered once, but all scenes were equally likely to appear in each of the face conditions over the course of the experiment.

Result

The data from one participant, whose search times were more than five standard deviations from the group mean, was excluded from all analysis. For the remaining 31 observers, accuracy, reaction times and search times are shown in Figure 6. A one-factor within-subject ANOVA showed a main effect of face type, $F(2,60) = 9.85$, $p < 0.05$, $\eta_p^2 = 0.25$. Tukey HSD test shows that this reflects reduced detection accuracy for vertically and horizontally stretched faces compared to their original counterparts, both $qs = 5.44$, $ps <$

0.001, $ds \geq 1.12$, while the two stretched conditions did not differ from each other, $q = 0.00$. $d = 0.00$.

A similar effect of face type was also found for response times, $F(2,60) = 26.63$, $p < 0.001$, $\eta_p^2 = 0.47$, and search times, $F(2,60) = 16.01$, $p < 0.001$, $\eta_p^2 = 0.35$. For both measures, Tukey HSD showed that the original faces were detected faster than their vertically and horizontally stretched counterparts, all $qs \geq 5.96$ $ps < 0.001$, $ds \geq 1.32$. In both response and search times, the two stretched conditions did not differ from each other, both $qs \leq 1.65$, $ds \leq 0.31$.

Discussion

The results of this experiment confirm that face detection is affected by vertical distortions and extend this finding to horizontally stretched faces. As in Experiment 2, this effect was found despite the fact that these stretched faces matched the surface area of their unstretched counterparts. This finding suggests that face detection relies on a template that utilizes typical height-to-width aspect ratios of faces. These findings are discussed in the General Discussion.

General Discussion

This study examined whether geometric distortions, by stretching faces to manipulate their natural height-to-width aspect ratio, impairs person detection. The impact of stretching on detection performance was not obvious when faces were equated to their original, unstretched counterparts in terms of their height or width dimension (Experiment 1). However, a clear effect of stretching was obtained when the original and distorted faces were matched for their surface area (Experiment 2), and this was found for both vertically and horizontally stretched faces (Experiment 3). This effect was evident in the accuracy and speed of observers' detection responses and also in the initial eye movements to faces, which

indicates that it arises during the *search* for faces in natural scenes. Moreover, this effect was found despite the fact that observers were informed of the stretched face conditions prior to the experiment. Taken together, these results suggest that the effect of stretching on face detection is remarkably robust.

These findings suggest that the height-to-width aspect ratio of faces is an important component of the cognitive template that is utilized for detection. Previous studies already suggest that this template might rely on a “quick and dirty” processing strategy that utilizes some salient but simple visual cues to locate likely face candidates. It has been shown, for example, that detection proceeds unhindered when internal (i.e., eyes, nose and mouth) or external facial features (e.g., face outline, hairstyle) are removed, as long as an oval face-shaped template is preserved (Hershler & Hochstein, 2005). Face detection is also facilitated by skin-colour tones but only when these are tied to the shape of a face (Bindemann & Burton, 2009). In contrast, detection performance is impaired when overall face-shape is destroyed by image scrambling (Hershler & Hochstein, 2005) or bit-part deletion (Burton & Bindemann, 2009). Taken together, these results indicate that face detection might be driven by a simple skin-coloured face-shape template. The current experiments add to these findings by suggesting that this template utilizes the natural height-to-width ratio of faces to aid detection.

To explore the role of such aspect ratios for face detection, the current study stretched faces vertically or horizontally to 200% of their original size, while maintaining the size of the orthogonal dimension. While this is a dramatic transformation, the question arises of whether the cognitive detection template is sensitive to smaller distortions that reflect natural between-subject variation of facial height-to-width ratios. To begin to explore this *a posteriori*, we calculated the response times to the original faces across all three experiments as a function of their height-to-width ratio. While these ratios ranged from 1.08 to 1.75, only

very few faces had such extreme ratios. We therefore divided the stimuli into larger non-overlapping face categories with height-to-width ratios that were close to 1.2, 1.4, 1.6 and 1.8. A one-factor ANOVA of this data, which is illustrated in Figure 7, showed an effect of ratio, $F(3,269) = 14.48, p < 0.01^1$, which reflects slower responses to faces in the 1.2 and 1.8 categories than for the two intermediate face ratios (Tukey HSD, all $ps < 0.01$). We obtained a similar pattern for search times, $F(3,241) = 3.45, p < 0.05$, which were slower for the 1.8 than the 1.6 and 1.4 categories (both $ps < 0.05$), while faces with a 1.2 ratio did not differ from any of the categories. Overall, these data therefore suggest that face detection is best with height-to-width ratios in the range of 1.4 to 1.6. We draw this conclusion tentatively, as these ratios were not manipulated systematically across our scenes.

The effect of geometric distortions on face detection is interesting considering that observers appear insensitive to subtle differences in the height-to-width ratio of individual face *identities* (Sandford & Burton, 2014), and as person *recognition* is also unaffected by the drastic manipulations that impaired the detection of faces in the current experiments (see, e.g., Bindemann, Burton, Leuthold, & Schweinberger, 2008; Hole, George, Eaves, & Rasek, 2002). This differential sensitivity to geometric distortions converges with other recent findings to indicate that detection differs from other tasks with faces (Bindemann & Lewis, 2013). In this respect, it is interesting to note that face detection might also differ from the perception of non-face stimuli, such as natural and urban scenes, which also appear to be insensitive to substantial linear distortions (e.g., up to 52%, see Kingdom, Field, & Olmos, 2007; see also Cutting, 1987).

¹ Some participants failed to record a single correct response in some of the height-to-width categories. Because of these missing data points, we computed ANOVA on a between-subjects basis.

References

- Bayliss, A. P., Pellegrino, G., & Tripper, S. P. (2004). Orienting of attention via observed eye gaze is head-centred. *Cognition*, *94*, B1-B10. doi:10.1016/j.cognition.2004.05.002
- Bindemann, M., & Burton, A. M. (2009). The role of color in human face detection. *Cognitive Science*, *33*, 1144-1156. doi:10.1111/j.1551-6709.2009.01035.x
- Bindemann, M., Burton, A. M., Leuthold, H., & Schweinberger, S. (2008). Brain potential correlates of face recognition: Geometric distortions and the N250r brain response to stimulus repetition. *Psychophysiology*, *45*, 535-544. doi:10.1111/j.1469-8986.2008.00663.x
- Bindemann, M., & Lewis, M. B. (2013). Face detection differs from categorization: Evidence from visual search in natural scenes. *Psychonomic Bulletin & Review*, *20*, 1140-1145. doi:10.3758/s13423-013-0445-9
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305-327. doi:10.1111/j.2044-8295.1986.tb02199.x
- Burton, A. M., & Bindemann, M. (2009). The role of view in human face detection. *Vision Research*, *49*, 2026-2036. doi:10.1016/j.visres.2009.05.012
- Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*, 361-380. doi:10.1111/j.2044-8295.1990.tb02367.x
- Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, *51*, 256-284. doi:10.1016/j.cogpsych.2005.06.003
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*, 286-291. doi:10.3758/BRM.42.1.286

- Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, *41*, 1179-1208.
doi:10.1016/S0042-6989(01)00002-5
- Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*, 641-651. doi:10.1038/nrn1724
- Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*, 985-994. doi:10.1068/p3335
- Crouzet, S. M., Kirchner, H. K., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10:4*, 1-17. doi:10.1167/10.4.16
- Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers in Psychology*, *342*, 1-9. doi:10.3389/fpsyg.2011.00342
- Cutting, J. E. (1987). Rigidity in cinema scenes from the front row, side aisle. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 323-334.
doi:10.1037/0096-1523.13.3.323
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Shared attention and the social brain: Gaze perception triggers automatic visuospatial orienting in adults. *Visual Cognition*, *6*, 509-540. Retrieved from <http://wexler.free.fr/library/files/driver>
- Farkas, L. G., Katic, M. J., Forrest, C. R., Alt, K. W., Bagic, I., et al. (2005). International anthropometric study of facial morphology in various ethnic groups/races. *Journal of Craniofacial Surgery*, *16*, 615-646. doi:10.1097/01.scs.0000171847.58031.9e
- Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*, 571-583. doi:10.1068/p5705
- Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, *45*, 1707-1724. doi:10.1016/j.visres.2004.12.021

- Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception, 31*, 1221-1240. doi:10.1068/p3252
- Jenkins, R. (2007). The lighter side of gaze perception. *Perception, 36*, 1266-1268. doi:10.1068/p5745
- Johnston, R. A., & Bindemann, M. (2013). Introduction to forensic face matching. *Applied Cognitive Psychology, 27*, 697-699. doi:10.1002/acp.2963
- Kingdom, F. A., Field, D. J., & Olmos, A. (2007). Does spatial invariance result from insensitivity to changes? *Journal of Vision, 7*, 1-13. doi:10.1167/7.14.11.
- Lewis, M. B., & Edmonds, A. J. (2003). Face detection: Mapping human performance. *Perception, 32*, 903-920. doi:10.1068/p5007
- Lewis, M. B., & Edmonds, A. J. (2005). Searching for faces in scrambled scenes. *Visual Cognition, 12*, 1309-1336. doi:10.1080/13506280444000535
- 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. doi:10.1002/ima.10040
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.
- Sandford, A., & Burton, A. M. (2014). Tolerance for distorted faces: Challenges to a configural processing account of familiar face recognition. *Cognition, 132*, 262-268. doi:10.1016/j.cognition.2014.04.005
- Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience, 31*, 411-437. doi:10.1146/annurev.neuro.30.051606.094238

FIGURE 1. Example stimuli for Experiment 1, depicting a scene without face (top left), and faces in the original (top right), horizontally compressed (bottom left), and vertically stretched condition (bottom right).



FIGURE 2. Detection accuracy (%), response times (ms), and search times (ms) for the face-present conditions in Experiment 1. Vertical bars represent the standard error of the means.

Face-absent trials: accuracy = 99.0% (SD = 0.1), response times = 1813 ms (SD = 124).

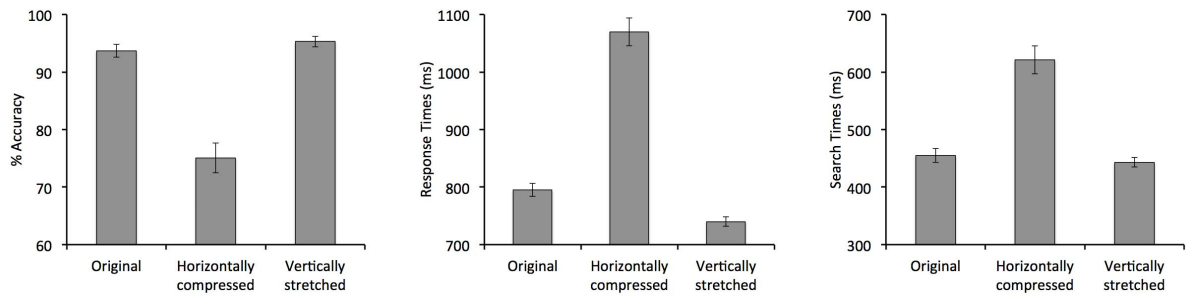


FIGURE 3. Example stimuli for Experiment 2, depicting faces in the original (top left), original large (top right), stretched (bottom left), and stretched large condition (bottom right).



FIGURE 4. Detection accuracy (%), response times (ms), and search times (ms) for the face-present conditions in Experiment 2. Vertical bars represent the standard error of the means.

Face-absent trials: accuracy = 99.0% (SD = 0.2), response times = 1666 ms (SD = 119).

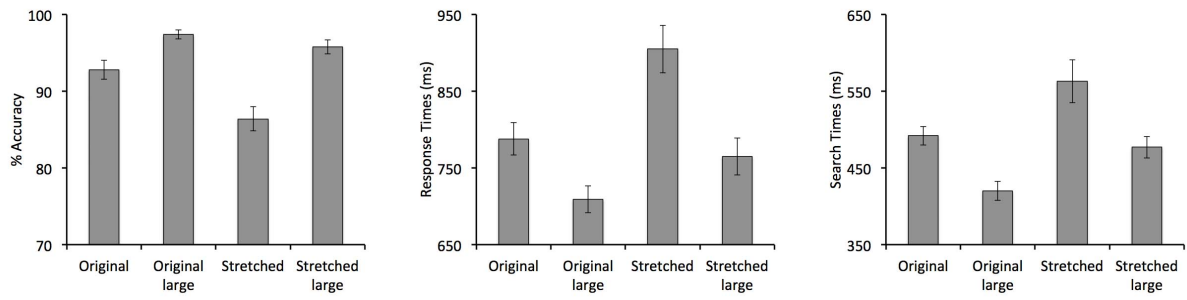


FIGURE 5. Example stimuli for Experiment 3, depicting faces in the original (left), horizontally stretched (centre), and vertically stretched condition (right).



FIGURE 6. Detection accuracy (%), response times (ms), and search times (ms) for the face-present conditions in Experiment 3. Vertical bars represent the standard error of the means.

Face-absent trials: accuracy = 99.0% (SD = 0.2), response times = 2007 ms (SD = 137).

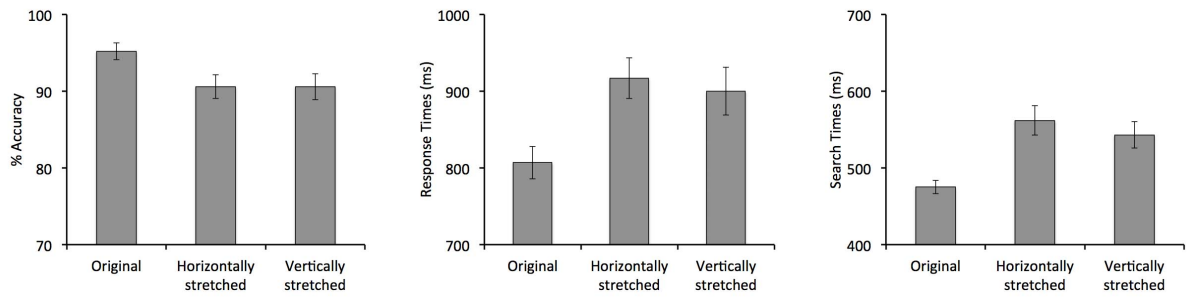


FIGURE 7. Response times (ms) and search times (ms) for the original face stimuli in Experiments 1 to 3, grouped by height-to-width ratio. Vertical bars represent the standard error of the means.

