Face detection: Mapping human performance

Michael B Lewis, Andrew J Edmonds
School of Psychology, Cardiff University, PO Box 901, Cardiff CF10 3YG, UK;
e-mail: lewismb@cardiff.ac.uk

Abstract. The recognition of faces has been the focus of an extensive body of research, whereas the preliminary and prerequisite task of detecting a face has received limited attention from psychologists. Four experiments are reported that address the question how we detect a face. Experiment 1 reveals that we use information from the scene to aid detection. In experiment 2 we investigated which features of a face speed the detection of faces. Experiment 3 revealed inversion effects and an interaction between the effects of blurring and reduction of contrast. In experiment 4 the sizes of effects of reversal of orientation, luminance, and hue were compared. Luminance was found to have the greatest effect on reaction time to detect faces. The results are interpreted as suggesting that face detection proceeds by a pre-attentive stage that identifies possible face regions, which is followed by a focused-attention stage that employs a deformable template. Comparisons are drawn with automatic face-detection systems.

1 Introduction

Humans are remarkably good at dealing with faces that they encounter. When presented with a complex visual scene, we are capable of locating a face, matching it to a known individual, and producing information about that person or a unique name. All this occurs effortlessly and, often, in well under a second. This ability has been the focus of considerable research activity over several decades. Consequently, we now understand how many aspects of this procedure may take place. Indeed, there exist detailed models to account for human performance on a variety of face-related tasks such as familiarity decisions, person identification, and name production. Recently, Burton et al (1999) offered a computational model to account for what they refer to as the complete process from 'pixel to person' (that is from the image to the accessing of knowledge concerning the person who is represented).

The focus of the research reported here concerns part of the process from pixel to person (or from image to identity), but a part that has not been incorporated in any of the main models of face recognition, from earlier ones (eg Hay and Young 1982), to the most recent (Burton et al 1999). All such models begin with the premise that a face is present and centred in the visual field. How a face is first detected and orientated towards remains poorly understood. The current research aims to elucidate how humans perform the initial process of face detection and, hence, move towards a complete model for face recognition that really does take us from pixel to person.

Before our research is described, what is already known about face detection is discussed. This background comes from two quite different sources: first from a limited cognitive-psychological literature; and, second, from extensive research into automatic-face-processing systems.

1.1 Previous psychological research

There has been little research conducted that focuses on how we are able to detect that an object is a face. There is, however, some literature on the differences between face recognition and object recognition, and a number of robust findings have been taken as evidence that different, even specialised, processes underlie face and object

recognition (eg Yin 1969; Valentine 1988; Tanaka and Farah 1993), or that faces are 'special' (eg Farah et al 1998). In view of what is still an ongoing debate, a pertinent question for this research is whether there are also differences in the detection of face and non-face objects. A number of ERP studies have addressed this question in recent years (eg Bentin et al 1996; Eimer 1998; Sagiv and Bentin 2001), suggesting that there is a face-specific encoding system, of which face detection is a part, which is distinct from mechanisms for object detection. There is little research, however, on behavioural differences, such as speed of reaction times and accuracy scores between face and object detection (although see Hill and Watt 1996, for an example). The current research addresses this, exploring detection of face objects across a number of tasks including detection from within a scene and the decision whether an image is a face under a variety of transformations.

Although there has been little published behavioural or psychological research into the factors affecting face detection by humans, there have been a number of tangentially related studies that might offer some insight into possible issues. This research can be divided into two types: face categorisation (eg face versus jumbled face decision) and analysis of the face-detection effect.

Valentine and Bruce (1986) conducted a series of experiments in which participants were presented with either faces or jumbled faces (ie the internal features of the face had been rearranged). The task of the participants was to indicate whether or not the image was an intact face. It could be said, therefore, that they were required to detect a face but in a limited situation (where the distractors could only be jumbled faces). Valentine and Bruce found that the task was performed faster for typical than distinctive faces and upright than inverted faces.

In a similar study, Cooper and Wojan (2000) presented participants with either faces, or faces that had one feature replaced by another (eg a mouth replaced by an eye). The task of the participants was to say if the image was a proper face or not. Some of the faces had one or two eyes raised from their natural position. Cooper and Wojan found that raising just one eye slowed this face categorisation task.

Visual-search paradigms have been employed in the task of judging the presence of an intact face (rather than a jumbled face). In such paradigms, the participant is required to indicate whether there is an intact face within an array of distractors made up of jumbled faces. Nothdurft (1993) employed such a visual-search paradigm with schematic faces. He found that, as the size of the array increased, the reaction time to detect the intact face increased. The conclusion drawn was that faces do not 'pop out'. That is, if the Treisman and Gelade (1980) interpretation is used, faces are detected with serial search not parallel search. Similar conclusions were drawn from experiments on real faces conducted by Brown et al (1997). These experiments also used arrays containing jumbled faces as distractors. In order for the task to evaluate more natural face detection (ie within a natural scene), the distractors should be stimuli that are more natural. More recent research conducted in parallel in two laboratories, however, showed that faces do in fact pop out (Elgavi-Hershler and Hochstein 2002; Lewis and Edmonds 2002). In these studies, the time taken to detect a real colour face among either non-facial objects or background-scene elements was tested. Lewis and Edmonds (submitted) further found that this pop-out effect was maintained when faces were inverted, reduced to grey-scale, or blurred. The evidence suggests, therefore, that in natural face detection faces are detected by a parallel pre-attentive search strategy.

The face-detection effect, as discovered and interpreted by Purcell and Stewart (1986, 1988, 1991), highlights one reason why face detection may be a special process and, therefore, requires focused examination. Their experiments involved finding the threshold presentation time for a stimulus to be detected. Surprisingly, they found that the threshold for a face was shorter than a visually matched image (eg an inverted

face or a jumbled face). This means that images that are upright faces are detected more easily than images that are not upright faces—an effect that was referred to as the face-detection effect. While these experiments do not reveal how faces are detected in complex scenes, they do reveal that the face has some unspecified special significance in the early processing of an image.

More recent work has supported this special significance of faces in that they tend to grab attention. Vuilleumier (2000), for example, found that patients with left spatial neglect showed more neglect when the competing object was a line-drawn face rather than a scrambled version and less neglect when the left-hand object was the face. The conclusions, therefore, were that attention is modulated by the stimulus and that faces 'call for attention'. A similar conclusion can be made from Ro et al's (2001) flicker-paradigm experiment. In this experiment a set of objects, including one face, was presented then removed and represented with one object changed. The change was detected faster when the object changed was a face—an effect that disappeared for inverted faces. This confirms the idea that faces automatically have an advantage when competing for attention.

Apart from this research, human face detection has largely been unstudied. Our research aims to rectify this with four experiments that explore factors affecting the speed of face detection. The results of these studies allow us to speculate how the process takes place. We also speculate how similar the processes of face detection by humans are to the strategies employed in automatic face-detection systems.

Computer scientists aiming to create automatic face-processing systems have found the issue of face detection almost impossible to ignore. Considerable effort has been focused upon finding ways to detect faces automatically in complex scenes. The ways that these scientists have found to overcome the problems of face detection may offer some insight into how the human face-detection system works to solve the same basic problem.

1.2 *Clues from automatic face detection*
In order to produce a fully automatic face-recognition system it is necessary to develop a system that can detect faces prior to processing identity. The problem of automatic face detection is one that has been considered in parallel to automatic face recognition. This latter area has provided useful insights into how humans may recognise faces. For example, the concept of an eigenface was developed from automatic face recognition but has been an important tool in understanding the psychology of face recognition (Turk and Pentland 1991). Automatic face detection, therefore, may similarly provide insights into understanding how humans detect faces.

While eigenfaces have been successfully employed for automatic face recognition, there is still no accurate and efficient method for detecting human faces (Han et al 2000). There have been, however, several methods proposed to do the task. These methods have employed a range of different features, that are available from faces, to overcome the problem of detection. The variety of these methods was reviewed recently by Yang et al (2001). Some of them are described here in brief.

A common and useful method of automatic face detection is to make use of the skin colour of faces. Lee et al (1996) made use of the fact that skin is never blue or green, in order to identify regions of an image that might be a face. Once such a region has been found, then a matching procedure can be performed to judge how similar that region is to a face template. A similar method of detecting possible face regions which they considered was to use dynamic information from moving images. These methods require specific information to be in the stimulus (either colour information or motion). Humans, however, can detect faces quickly in static black-and-white images.

While colour and movement might be employed by the human face-detection system, they are not necessary for the task.

It is possible for automatic systems to detect faces from static grey-scale images. Juell and Marsh (1996) showed this using edge detection to identify facial features and a back propagation network to identify the faces within the scene. This system, like many others, only worked if faces were upright and had frontal views. Generalising a system to recognise a face in any view is considerably harder, but a possible solution was offered by Han et al (2000).

Han et al (2000) used the fact that facial features (eyes and mouth) are usually indicated by an area of high local contrast within an area of low contrast (skin areas). Their system identifies these local high-contrast areas and marks them. In order to find faces, the system examines each pair of local high-contrast areas and attempts to fit a standard face template to it such that the high-contrast areas are the eyes of that face. There are, of course, standardisation procedures and tolerance built into the system so that even partially occluded faces can be detected. The resulting system can find faces efficiently and fairly reliably from a grey-scale static and cluttered image (even when faces are rotated or inverted).

Automatic face-detection systems offer a range of possibilities for how human face detection may take place, only some of which have been discussed here. It is quite possible that the human face-detection system makes use of all or some of these techniques. Alternatively, human face detection may employ methods not yet considered by computer scientists. The experiments described below address this issue.

2 Experiment 1: Faces in scenes

Automatic face-recognition systems, in general, do not make use of the non-facial information contained in a scene that may be helpful for identifying the location of possible faces. A face, therefore, would be just as difficult to detect if it was on top of a tree as on top of a human body. This failure to use scenic information to identify possible locations for faces is probably due to the fact that it would be harder for an automatic system to recognise a human body than to detect a face and so no computational advantage can be gained by using scenic information.

The human brain, on the other hand, may be able to process scenes in parallel and so scenic information could be processed whilst detection of a face continues. This might mean that it is possible that information from the scene can be used to guide the detection of faces. In this first experiment a face-detection task is used to investigate whether scenic information plays a role in natural face processing. If we do indeed use information from the surrounding context to help us to detect the presence of a face, then we should observe faster face detection when a face occurs in its natural position than when a face occurs in a scrambled scene.

Two further transformations were investigated. The first transformation considered was inversion. It has been shown previously that inverting a face makes it harder to recognise (eg Yin 1969) possibly because configural encoding is disrupted (Tanaka and Farah 1993). Inversion has been found to slow face-classification tasks (ie faces versus jumbled faces—Valentine and Bruce 1986) which may be similar to face-detection tasks. If inversion of a face slows face detection then this may imply that configural encoding is employed in that task also. Maurer et al (2002), however, consider that there may be various aspects of configural encoding: first order, holistic, and second order. These aspects may be differentially affected by inversion, but also their roles may be different. They suggest that first-order configuration may be important for face detection, whereas holistic and second-order are important for recognition. Inversion effects, therefore, are not necessarily the same for detection as they are for recognition.

Inverting a face is thought to disrupt some aspect of configural encoding but leave featural encoding intact. Conversely, so it has been argued (Collishaw and Hole 2000), blurring disrupts featural encoding while leaving configural encoding intact. Featural encoding is the piecemeal analysis of a face that is slow but unaffected by inversion (see Tanaka and Farah 1993). If small amounts of blurring are found to slow face detection, then this would suggest that featural encoding is used in face detection. The interesting analysis, however, is how inversion and blurring interact. Human face processing is resourceful particularly under sub-optimal conditions. Face detection may proceed by featural and/or configural encoding of the image. If one system of encoding is unavailable, then the other system will be relied upon. A direct consequence of this is that there would be a greater decrement in recognition for blurred inverted faces than would be expected by adding together the decrement for either transformation on their own. Such a set of results was found in a face-recognition task (Collishaw and Hole 2000).

In the current experiment we explored the effects of inversion and blurring as well as presenting the image within a scene on a face-detection task. We will return to these first two effects in latter experiments but in the current experiment they are important because they show how the three factors all interact within the process of face detection.

2.1 Methods

2.1.1 *Participants.* Twenty-one students received course credit for their participation. All had normal or corrected-to-normal vision.

2.1.2 *Stimuli.* All stimuli were generated from colour digital-video captured images from a popular UK soap opera. Eighteen scenes were selected that contained faces and bodies of different characters. The faces within each image took up approximately 100 (hairline to chin) by 80 (face width) pixels. The images were presented at 72 pixels per inch and were viewed at a distance of approximately 0.5 m. Each image was cropped to 360 by 289 pixels. A black 3×3 rectangular grid was superimposed on the picture to tessellate it into regular cells (the internal size of each cell was 114 by 90 pixels). The face in the image fell wholly into one of the top three cells of the image. An example of the stimuli is shown in figure 1. The images generated in this way were used as the basic untransformed stimuli.

Further stimuli were generated by performing a number of transformations to the stimuli. The first transformation was removal of the face cell. A cell from the bottom of the image was duplicated and pasted to cover the face cell. The second transformation was scrambling of the cells. Within each of the three rows of the image, the three cells were rearranged such that they occurred in a different order from the original, a different order from the other two rows, and no two vertically adjacent cells remained vertically adjacent. The third transformation was inversion. The whole image was inverted. The fourth transformation was blurring. Corel Photopaint was used to perform a Gaussian blur (radius set to 3 pixels).

The four transformations were performed factorially in order to generate 16 stimuli for each of the eighteen original scenes. The entire set of 288 stimuli was employed in the experiment.

2.1.3 *Procedure.* Participants were sat approximately 0.5 m from a computer monitor. Each of the 288 stimuli was presented one at a time in a random order. The task of the participants was to make a speeded decision whether each image contained a face or not. Participants made their decision by pressing one of two keys on the keypad. Once the participant made his or her response, the image disappeared and the next image appeared after an interval of 1 s.

Figure 1. Examples of the stimuli employed in experiment 1. Top-left: upright, intact, and unblurred. Top-right: upright, scrambled, and unblurred. Bottom-left: upright, intact, and blurred. Bottom-right: upright, scrambled, and blurred. Invert the figure to see examples of inverted stimuli. Stimuli were presented in colour.

2.2 Results

The data from the positive (face-present) trials and negative (face-absent) trials were analysed separately. Errors and trials with reaction times greater than 100 ms were removed from the calculation of participant means (this accounted for less than 5% of the data). The cell means for the data are shown in figure 2. It was found that the time to respond that a face was present in the scene was increased by scrambling the scene, inverting the image, and blurring the image. These effects appear to have very little interaction between them. For the face-absent trials, orientation had very little effect on reaction times; however, scrambling the scene and blurring each appeared to speed correct negative responding.

A three-way ANOVA was performed on the data for face-present trials. This found a significant effect of scrambling ($F_{1,20} = 48.358$, $p < 0.001$), inversion ($F_{1,20} = 68.441$, $p < 0.001$), and blurring ($F_{1,20} = 93.966$, $p < 0.001$). The two-way and three-way interactions were all non-significant ($Fs_{1,20} < 3$).

A three-way ANOVA was performed on the data for face-absent trials. This revealed a significant effect of scrambling ($F_{1,20} = 10.031$, $p < 0.01$), and blurring ($F_{1,20} = 39.114$, $p < 0.001$). Inversion had a non-significant effect on reaction times ($F_{1,20} < 1$). The interaction between blurring and scrambling was also significant ($F_{1,20} = 13.147$, $p < 0.001$). Further analysis revealed that there was an effect of blurring for scrambled and non-scrambled images but the effect of scrambling was only significant
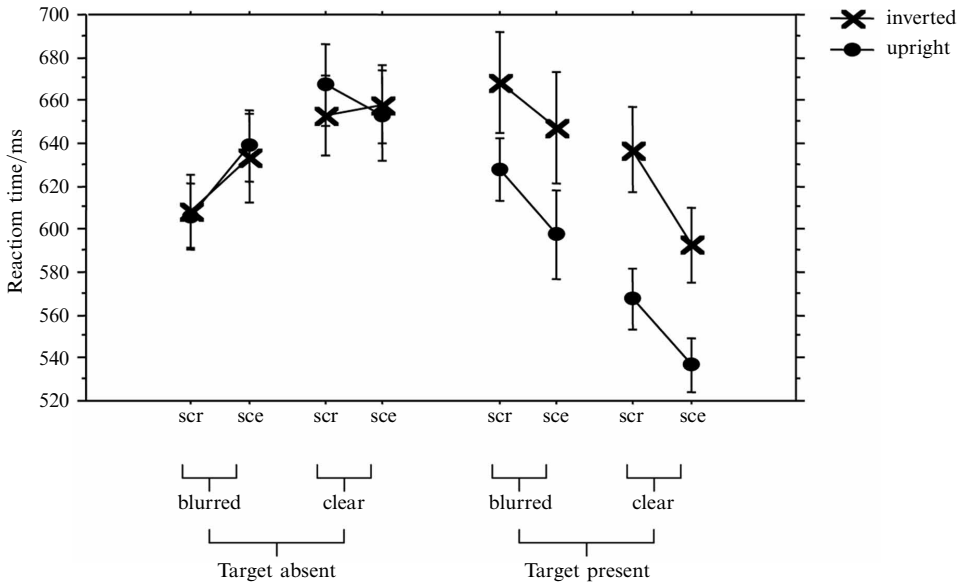
Figure 2. Mean reaction times to make a correct face-present or face-absent response for the different conditions of experiment 1. scr = scrambled scene; sce = intact scene. Error bars show 95% confidence intervals based on within-participants variability (see Loftus and Masson 1994).

for blurred images. The other two two-way and the three-way interactions were all non-significant ($Fs_{1,20} < 3$).

2.3 Discussion
The most important finding of this experiment is that a face presented in a natural scene is detected faster than the same face where the surrounding scene has been scrambled. A possible explanation of this result is that we use scenic information from the surrounding image in order to help detect a face. This surrounding information may help to direct attention towards an area in the scene that is likely to contain a face (eg above the shoulders). There are, however, at least two alternative explanations. First, this surrounding information may speed detection by contributing towards a confirmatory process that speeds the decision that what is attended to is a face and not anything else. Second, it may be the case that scrambling any scene makes detection of an element harder. Disruption of the context will have an inhibitory effect.

Evidence from the target-absent trials, however, offers some support for the directed-attention explanation. In these trials, scrambling the scene actually speeds a correct negative response (although only for the blurred scenes—the clear scenes show no significant differences). If the scenic information is directing attention to where the face should be and it is not there, then more time will be required to search the rest of the scene than if attention was spread equally across the scene. This pattern of data, therefore, suggests that scenic information has its influence in a pre-attentive manner that directs attention to find a face. Such a system would be behaviourally useful, as typically a scene would provide valid information where to locate a face. The current data, however, cannot necessarily be used to discount a role for an inhibitory bottom–up effect of scrambling the scene.

Second, a blurred face takes longer to detect than a non-blurred face. Blurring of a face obviously removes fine detail of the image and this removal of information means that it would be surprising if it had no effect on fact detection. What this result does tell us, however, is that human face detection is, indeed, influenced by fine-detail information—ie it is not exclusively a coarse-detail procedure as selection based on

colour, for example, might be. Blurring, however, had an inverse effect on the time required to state that no face was present. In fact, in the case of the blurred and scrambled scenes, the time for a negative response was shorter than the time required for a positive response. An explanation for this pattern of results is that, in the target-absent trials, blurring acts to reduce the amount of visual information that needs to be searched, whereas, in the target-present trials, it acts to reduce the confidence in making a response.

Third, inversion of the image increases the time to decide that a face is present. Unfortunately, not much can be made of this result, as there is a confound: faces that were inverted were all presented in the bottom three cells of the image, whereas upright faces were presented in the top three cells. This observed inversion effect, therefore, could be an effect of face inversion or face location. Further research is required to distinguish these two possible explanations.

Fourth, the three effects of scrambling, inversion, and blurring appear to be approximately additive as regards their influence on reaction times. This means, for example, that the size of the effect of scrambling is the same regardless of whether the face is upright or inverted, or blurred or not. It is possible, therefore, that the effects of scrambling, inverting, and blurring are affecting three different aspects of the process of face detection. This interpretation, however, relies on the application of the additive-factors method (Sternberg 1969) and so is tentative. There are problems arising with this method which are highlighted elsewhere (eg McClelland 1979; Miller et al 1995; Lewis et al 2001). If this interpretation is to be believed, however, then human face detection is a complicated, multistaged process. Even without the application of the additive-factors method, the results still indicate that there is a range of effects that need to be explained in order to understand how face detection takes place.

One aim of this research is to offer a description of how humans detect faces in order to inform us how an automatic system might perform a similar task. This first experiment has demonstrated that, possibly, orientation is important in human face detection. There already exist automatic methods of face detection that are orientation-dependent. These methods tend to be the template-driven methods. Also, blurring of a face is likely to reduce the ability of almost any automatic face-detection system because they tend to rely on detailed information as well as, sometimes, coarse information. Human ability, therefore, reflects the likely performance of automatic systems when faces are blurred. Automatic face-detection systems, however, do not tend to make use of non-facial information, either to reduce the search space or to improve the accuracy of the search. Humans, it would appear, do use information from the surrounding scene in order to speed face detection. If one wants to build an automatic system that performs like a human, therefore, one will have to build a system that analyses the entire image for extra-facial as well as facial information.

3 Experiment 2: Use of specific features

Experiment 1 indicates that face detection is guided in a pre-attentive manner from scenic information and features of the face itself. Further research to find how the process of face detection takes place once attention has been guided, therefore, will require analysis of simple scenes that either do or do not contain a face. By presenting just the cells that were used in the previous experiments in isolation, these next experiments explore what aspects of the face affect the time required to determine that it is indeed a face, once attention has been directed towards it. The remaining three experiments use a similar isolated-face-detection paradigm to investigate different aspects of the task. In the first of these, we investigate the relative contributions made by different parts of the face towards rapid detection.

Several of the automatic face-detection systems described above employ specific features as part of the overall process (eg Han et al 2000). One of the most consistent and salient aspects of a face is the region around the eyes, as has been demonstrated in the batman effect (eg O'Donnell and Bruce 2000). Covering the eye region makes recognition of a person very difficult.

In experiment 2, we explore whether the eyes play a special role in the detection of faces by humans. By placing a rectangle across the eye region it was possible to determine whether obscuring the eyes affected face detection. This was contrasted with placing a similar rectangle across other features. In this way, the relative importance of the eye region, compared to the mouth region for example, could be determined.

3.1 Method
3.1.1 *Participants*. A further thirty undergraduate students received course credit for their participation in the experiment. All had normal or corrected-to-normal vision.

3.1.2 *Design*. The experiment was a within-participants design with seven levels (no face, whole face, forehead obscured, eyes obscured, nose obscured, mouth obscured, and chin obscured). The dependent variable was the time required to decide whether the image was of a face or not.

3.1.3 *Stimuli*. In this experiment, the single face cells described in experiment 1 were used as the positive (face) stimuli. Each of the eighteen faces was shown in each of the six 'feature obscured' conditions, making a total of 108 positive stimuli. The six conditions were either the unaltered face or the face with a black rectangle over it at one of five locations: the forehead, the eyes, the bottom of the nose, the mouth, or the tip of the chin. Figure 3 illustrates the nature of the stimuli. A further 162 images made up the non-face cells. These cells were taken from the non-facial background cells that made up the stimuli in experiment 1. Eighteen of these were not obscured, whilst the remainder were obscured in the same way as the face cells, ie by a black rectangle at a variety of locations.

Figure 3. Examples of the types of stimuli employed in experiment 4. From left to right the images show: forehead obscured, eyes obscured, nose obscured, mouth obscured, and chin obscured.

3.1.4 *Procedure*. Participants were told that a single image would appear in the middle of the screen, and that their task was simply to decide whether the image was of a face or not, via a key-press, as quickly and accurately as possible. Each image remained on the screen until a response had been given.

3.2 Results
All errors and reaction times over 1000 ms were discounted prior to the analysis, which led to the exclusion of 6% and 3% of all items, respectively. The mean reaction time to decide whether an image was a face or not is shown in figure 4.

A one-way within-participants ANOVA showed a highly significant main effect of face ($F_{1,29} = 31.339$, $p < 0.001$). Thus, participants took significantly longer to correctly
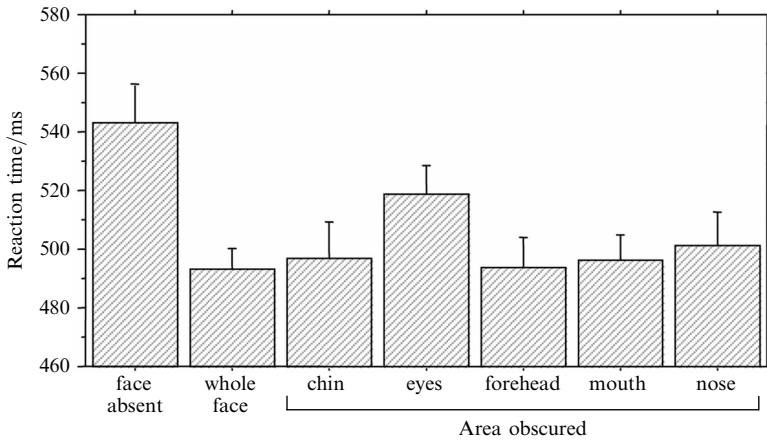
Figure 4. Average reaction times to make a correct face versus non-face decision in experiment 2. Error bars show 95% confidence intervals based on within-participants variability (see Loftus and Masson 1994).

decide that an image was not a face than to correctly decide that it was. A one-way within-participants ANOVA of the face-present data showed a significant main effect of obscuring the features of the face ($F_{5, 145} = 5.132$, $p < 0.001$). A series of planned pairwise comparisons revealed that covering the eyes of the person produced significantly longer reaction times compared with the effect of covering each of the other features, and with seeing the whole face (eyes covered versus nose covered, $p < 0.01$; in all other cases $p < 0.001$). Thus it would appear that covering or obscuring the eyes is particularly detrimental to the process of face detection.

3.3 Discussion

The results of experiment 2 clearly demonstrate that detection of a face is affected by obscuring the eyes of the face. While reaction times also increase when other features are obscured, these effects do not reach significance and are less than the effect of obscuring the eyes. The conclusion that must be drawn, therefore, is that the eyes form an important part of the pattern for face detection.

There are several ways in which the eyes may be important in the detection of faces. The first is that the eyes contain a great deal of variation in the changes from light to dark. The quality of the images employed here, however, was not of a sufficiently high resolution to pick up the finer spatial details of the eyes. Indeed, there is a similar amount of luminance variability within the mouth region as within the eye region in the images employed here. This effect may still be larger than the effect for the mouth because, in general, a face contains two eyes. This point is fundamental to the second reason for the special role of the eyes. The eyes could be extracted quickly from an image as a pair of (approximately equally sized) darker regions arranged at a similar height. This description would provide an accurate description of regions that are probably eyes and if they are not eyes then they might look eye-like. This would be consistent with Cooper and Wojan's (2000) finding that raising one eye above another slows a face-categorisation task. A third way in which eyes may be specifically important for face detection is that they form part of some larger configuration that aids detection. It may in fact be the relationship between eyes and the other features that provide their special role in face detection.

4 Experiment 3: Inversion, blurring, and contrast

In experiment 2, we investigated specific facial features, whereas in the remaining two experiments we explored changes to the whole image. These changes were made by a number of different global transformations. As described above, the effects of inversion and blurring have been important in the study of face recognition. These effects may also be important for face detection and experiment 1 offers some support for this.

A third factor considered in this experiment is the role played by contrast. Some automatic face-detection systems operate by locating areas of high contrast or areas of low luminance in order to identify candidate areas for faces to which a template is compared (eg Han et al 2000). If the human face-detection process uses a similar method, then it would be expected that face detection would be more efficient for high-contrast images than for low-contrast images. We consider this by investigating the effect of reducing the visual contrast of the images as well as inverting and blurring them.

4.1 Method

4.1.1 *Participants.* Twenty-four participants received course credit for taking part. All had normal or corrected-to-normal vision and none had taken part in experiments 1 or 2. The participants were taken from Cardiff University participant panel and had a mean age of 20 years.

4.1.2 *Stimuli.* The stimuli were constructed from the cells generated from the video-captured images used in experiment 1. These basic images were converted into 256-level grey-scale. These images were transformed in three ways factorially to generate eight types of images. First, they were inverted. Second, Gaussian blur was applied (as described in experiment 1) with a radius of 3 pixels. Third, the contrast was reduced by 50%. All the face cells and half of the non-face cells were transformed in these ways to give 720 stimuli in total (from the eighteen original scenes).

4.1.3 *Procedure.* Participants were sat approximately 0.5 m from a computer monitor. Each trial began with the presentation of a cell, which either contained a face or did not contain a face. The task of the participants was to press a key if the image contained a face. If no key was pressed then the trial ended after 800 ms and responses made after this time were recorded as errors. There was a 1 s interval before the next trial began. The order of presentation of the trials was randomised individually for each participant.

4.2 Results

The mean reaction time to determine that an image is a face was collected for each participant in each of the eight conditions. These means are shown in figure 5. Each of the three transformations increased the reaction time for the decision that the image was a face.

A three-way factorial ANOVA was performed on the data with within-participants variables of three possible transformations. This analysis showed simple effects for each of the three transformations: inversion ($F_{1,23} = 46.070$, $p < 0.001$); blurring ($F_{1,23} = 181.888$, $p < 0.001$); and contrast change ($F_{1,23} = 186.914$, $p < 0.001$). Only one interaction was significant, that between contrast and blurring ($F_{1,23} = 55.091$, $p < 0.001$). Simple-main-effects analyses confirmed that contrast had significant effects at each level of blurring and vice versa. The simple-main-effects analysis confirms that the interactions represent a change in the size of the effects across conditions and not the absence of effects in some conditions.

4.3 Discussion

The three manipulations of the image each had a significant effect on the speed of face detection: inverted faces took longer to detect than upright faces; low-contrast faces took longer to detect than high-contrast faces; and blurred faces took longer to detect than
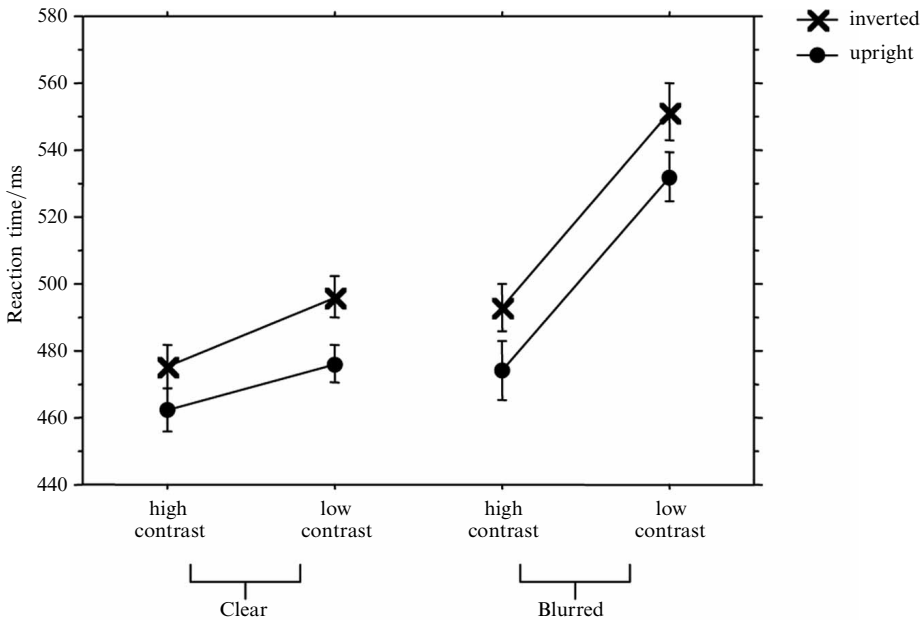
Figure 5. Mean reaction times to make a correct face-present response for the different conditions of experiment 3. Error bars show 95% confidence intervals based on within-participants variability (see Loftus and Masson 1994).

clear faces. Two of these effects confirm that the effects seen in experiment 1, of inversion and blurring, were not wholly due to scene effects. These two effects, together with the contrast effect, begin to inform us about what aspects of the visual form are important for face detection. Inversion of a face makes face detection slower by about 17 ms across all conditions. Reducing the contrast or reducing the clarity of the image also slow face detection but they do so in a super-additive manner such that their combined effect is larger than the sum of the simple effects of blurring and contrast reduction.

It is tempting to interpret the reaction-time data in a stages-of-processing manner (cf Sternberg 1969). The universal and constant effect of inversion implies that this affects some mandatory stage of processing and does so independently of other aspects of processing. Such a stage could be rotation of the image to compare it to some vertically orientated face template. If this is the case, then one would expect a gradual change in face-detection speed as a face is rotated from upright to inversed. Such a prediction may be tested in future research. The super-additive effects of clarity and contrast may mean that they both affect the same stage of processing. A possible candidate for such a stage would be edge extraction. Blurring and reducing contrast would each make it more difficult for edges to be found in the image. If edge extraction is an important process in face detection, then this is the pattern of results we would expect. An alternative stage could be a comparison between the image and a face-luminosity template. Blurring and contrast reduction would each reduce the degree of correlation between the image and a face template. The super-additivity of contrast and blurring could be accounted for in other ways. It is possible that there is more than one route to a face being detected and some of these routes are differentially affected by blurring and contrast reduction. As these routes are complementary, then disruption to either will not have the same effect as disruption to both. In many ways, these possibilities (a single stage or two parallel stages) are similar. They both place the loci of the effects of blurring and contrast reduction at a single stage that is sequential to the locus of effect of inversion.

It is interesting that no significant interaction was found between blurring and inversion. Collishaw and Hole (2000) have shown in a face-recognition task that blurring and inversion have a super-additive effect in their disruption of identity processing. It is thought that, as inversion disrupts configural encoding, and blurring disrupts featural encoding, the two forms of disruption together make it almost impossible to retrieve identity from an inverted and blurred face. The lack of an interaction between inversion and blurring for face detection ($F_{1,51} < 1$) means that the same kind of partition of information between configural and featural encoding is not appropriate in this task. A conclusion of this experiment is therefore that the information required to detect a face is different from the information required to recognise a face. The fact that face detection is not as heavily dependent on configural encoding as is face identification also accounts for the relatively small size of the inversion effect (only 17 ms). Valentine (1991) reported that the effect of inversion on the reaction time for a familiarity decision was 902 ms and even for a face-categorisation decision (face versus non-face), it was 129 ms. The configural information that appears to be so important for face identification seems to have a minor (but significant) effect on face detection.

5 Experiment 4: Inversion, luminance, and hue

The colour of a face has provided a useful tool for systems attempting to detect faces. Although faces do vary considerably in their colour, they are never naturally blue or green. Even across races, skin colours are remarkably consistent (Wu et al 1999). A face-detection system such as that proposed by Lee et al (1996) can employ colour to help identify possible locations in which to look for a face. If such a system is effective for automatic face detection then it might also be employed by humans in doing the same task. It is obvious (and it was confirmed in experiment 3) that faces can be detected in the absence of colour, but this does not necessarily mean that colour is of no importance when it is present.

While colour might be useful for face-detection systems, patterns of light and dark almost certainly are (after all, patterns of light and dark are all that is left in the grey-scale images of experiment 3). How we make use of this information, however, is unknown. Do we, for example, look for areas of high contrast within areas of low contrast as the face-detection system by Han et al (2000) works (ie eyes within a face)? Alternatively, do we match a luminosity template to possible facial portions of an image (cf PCA systems described by Gong et al 2000)? While there are many other possibilities that are not yet considered, these two can be distinguished by exploring the effect of luminance reversal. If it is the pattern of contrast that is important, then luminance reversal should not affect the detection of faces. Luminance reversal, however, would have a considerable effect on the ability of a process that involved template matching.

In the following experiment we explore the relative effects of hue reversal and luminance reversal on a simple face-detection experiment. The effect of inversion is also explored in order to compare its size against the size of the other two effects. Such comparison can provide further information about the nature of information required to make a face-detection decision by comparing the task with other face-processing tasks. For example, the matching of faces or Thatcherised faces (a task that clearly involves configural encoding) is slowed more by inversion than negation (Lewis and Johnston 1997). Similar studies were conducted by Hole et al (1999) who investigated the chimera effect for inverted and negated faces. They found that the hindrance to matching the top halves of chimeric faces was reduced by inversion but not negation. Inversion and not negation, therefore, interferes with configural encoding. Kemp et al (1990) found that, in a feature-displacement task, performance was impaired by negation but impaired to a greater degree by inversion. Johnston et al (1992)

compared the relative effects of inversion and brightness reversal (with grey-scale faces). Using the recognition parameter A', they found that inversion and negation affected recognition to approximately the same degree (the effect of inversion was slightly larger) and these effects had an apparently additive effect. They concluded, therefore, that while inversion disrupts the holistic encoding of a face (synonymous with configural encoding) negation disrupts the extraction of shape from shading.

The following experiment contrasts the effects of inversion, luminance reversal, and hue reversal on the speed of face detection. If configural encoding is more important for face recognition than for face detection then we would predict a smaller effect size than for luminance reversal. The hue variable enabled the relative importance of colour in face detection to be analysed.

5.1 Method

5.1.1 *Participants.* Twenty-eight students received course credit for their participation or received a small payment. All had normal or corrected-to-normal vision and none had taken part in experiments 1, 2, or 3.

5.1.2 *Stimuli.* The stimuli were constructed from the cells generated from the video-captured images as in experiment 3 but this time the images were kept in colour rather than being converted into grey-scale. The images were adjusted, however, to standardise their luminance histograms. These resulting images were transformed in three ways in order to generate eight types of factorial images. First, they were inverted as in experiment 3. Second, the pattern of luminance of the image was reversed such that bright areas became dark areas. Third, the pattern of the hue was rotated such that colours took their opposite values on the colour wheel. These transformations were applied to all the face and non-face cells alike. All the face cells and half of the non-face cells were transformed in these ways making 720 stimuli in total (from the eighteen original scenes).

5.1.3 *Procedure.* The procedure was identical to that employed for experiment 3. Participants had to make speeded face-detection decisions while each of the 720 stimuli were presented for 800 ms with a 1000 ms interval between trials.

5.2 Results

The mean reaction times to determine that an image was a face were collected for each participant in each of the eight conditions. These means are shown in figure 6. Each of the three transformations increased the reaction time for the decision that the image was a face.

An ANOVA was performed on the data with the three possible transformations as factors. This analysis found simple effects for: inversion (effect size $= 16$ ms; $F_{1,27} = 46.059$, $p < 0.001$); luminance reversal (effect size $= 35$ ms; $F_{1,27} = 202.046$, $p < 0.001$); and hue reversal (effect size $= 8$ ms; $F_{1,27} = 9.377$, $p < 0.01$). None of the interactions reached significance ($Fs_{1,27} < 3$).

As inversion, hue reversal, and luminance reversal are all dichotomous changes, it is meaningful to contrast the size each effect has compared to the other. To avoid any confounds from interactions (which may be there, although they are non-significant) a one-way ANOVA was conducted with four levels: untransformed, inverted, hue reversed, and luminance reversed. Only one change had been made to the last three sets, whereas the first set acted as a control. The ANOVA was significant ($F_{3,81} = 19.452$, $p < 0.001$). Planned pairwise comparisons revealed that there were significant differences between the control condition (untransformed) and the inverted and luminance-reversed conditions. The inverted condition was, in fact, significantly different from any of the other three conditions (ie the decrement due to inversion was significantly greater than the decrements due to either luminance reversal or hue reversal). All other comparisons were non-significant.
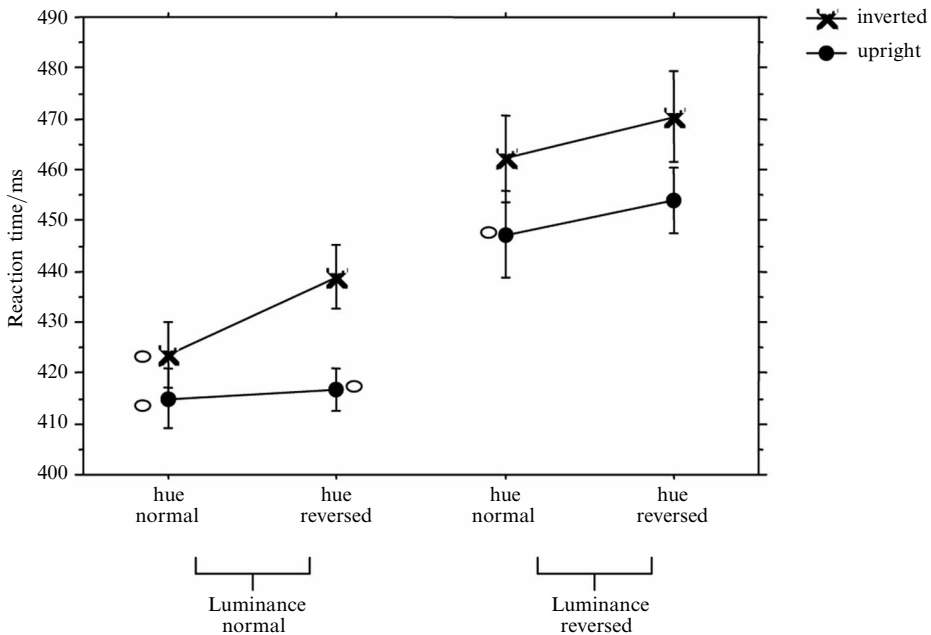
Figure 6. Mean reaction times to make a correct face-present response for the different conditions of experiment 4. Error bars show 95% confidence intervals based on within-participants variability (see Loftus and Masson 1994). The cells with the open circles next to them indicate those that were employed for the one-way ANOVA.

5.3 Discussion

Experiment 4 demonstrates that inversion, hue reversal, and luminance reversal each slow the process of face detection. Further, these three effects act in a roughly additive manner. One account, therefore, is that these additive effects may be influencing different stages of the process of face detection. There are, of course, other possible ways that the factors may influence the face-detection process and provide this pattern of results.

The important results from this experiment come from the comparisons of the relative size of the simple effects. The order of the effects in terms of size is luminance as the greatest, followed by inversion, followed by hue. The order of the first two effects is in contrast to the order found in face-recognition studies (eg Johnston et al 1992) or tasks based on configural encoding (eg Hole et al 1999; Lewis and Johnston 1997). On such tasks it has been found that inversion has at least the same, if not a greater, effect than negation. The fact that luminance has a greater effect than inversion on the face-detection tasks means that there is something quite different about this task when compared with other face-related tasks. This difference may exist for any number of reasons, but it is likely that configural encoding is not as important for this task as it is for the other tasks discussed.

The fact that the effect of luminance is much greater than the effect of hue implies that, if colour is used to speed face detection, then it has a minor role to play. If colour does help to orient attention to the face, then the benefit found for this process in the current paradigm is small when compared to that found for other factors. In the current experiment, it could be that colour acts to increase confidence of responding once the face has been detected. Colour is, therefore, just one of many pieces of information that come together in order to complete a face-detection task.

6 General discussion

These four experiments do begin to offer some insight into the task of face detection but they also throw up as many questions as they answer. Here, it is discussed what has been learnt about the process of face detection as well as some of the important issues that remain.

The most important contribution of experiment 1 is that it revealed that the context in which a face is viewed does affect the speed of face detection. A face in a natural scene (ie where it appears on top of a pair of shoulders) is detected more quickly than a face in a scrambled scene. This effect is important, because it immediately reveals that human face detection does not follow any of the computer methods exclusively. None of the automatic face-detection systems described above would find a face faster (or more accurately) in a scene rather than a scrambled scene. The evidence suggests that we use information from the scene in a pre-attentive or early-attentive manner to direct attention to regions that are likely to contain faces. This first experiment raises many questions how the scenic information may be processed in order to direct attention towards a face. It implies that in order to fully understand face detection we will also have to understand scene analysis.

In experiment 2, we investigated the role of isolated features in the task of face detection. The feature that affected face detection the most when obscured was the eyes. This highlights the special role played by the eyes in face processing. The faces could still be detected even when the eyes were removed, which indicates that there must be more than one strategy for face detection: one which is fast and involves locating the eyes; and a second which is slower and employs other features.

Experiments 3 and 4 each showed that inversion affected the face-detection decision for a simple image. This effect of inversion did not interact with any of the other four remaining factors. The constancy of the size of the inversion effect across all conditions suggests that it affects face detection in a unique way. It is possible that this stage requires the reorientation of some face template in order to make a match. Alternatively, there may be two templates—one for an upright face and one for an inverted face—but the template for the upright face would be faster to activate because it is used more often. Of course, other explanations for the inversion effects may be possible but are not considered here. The size of the inversion effect is also important, particularly when compared to the size of the luminance effect in experiment 4. Inverting an image slows face detection but not to the same extent as reversing the luminance of the image. Conversely, in many face-processing tasks (such as detecting Thatcherisation—Lewis and Johnston 1997), inversion is more disruptive than luminance reversal. This difference suggests that the inversion effect seen for face detection is qualitatively different from that seen for recognition. It is generally accepted that inversion disrupts recognition because of disruption to configural encoding. It appears from the data of experiment 4 that the same kind of account of inversion effects is not appropriate for the face-detection task. The inverted-template account would appear to be the better account.

The fitting-to-template account of focused face-detection decision can also help to understand the remaining results of experiments 3 and 4. These experiments revealed that reducing contrast and blurring slowed detection in a superadditive manner, luminance reversal had a large effect, but hue reversal had a small (but significant) effect. The first pair of results fits with the comparison of the image to some standard template. Reducing contrast of the image would obviously reduce the degree of similarity between the image and the template (if it is assumed that there is also some constant noise in the system). Blurring the image would have the effect of removing high levels of contrast variability, and so blurring would reduce the contrast variability between the template and the image. The fact that these effects are superadditive

further supports this template-matching view (by application of additive-factors method). Luminance reversal has the largest effect on face-detection speed (although direct comparison with blurring and contrast are meaningless, as these changes were arbitrary). If the template being employed is stored as a luminance map, as is common in the automatic-face-detection literature, then one can see why luminance reversal is so disruptive to face detection. Reversal of the luminance will lead to a complete lack of fit to the template. One way around this lack of fit would be to employ templates that encode changes in luminance rather than absolute luminance patterns (effectively edge detectors). It is possible that such edge detecting is going on, which allows the luminance-reversed face to still be detected albeit much more slowly than a normal face. Hue plays a small role in face detection. This may be because the face-detection templates rarely use hue or because they do not use hue at all but it still provides a useful confirmation of detection.

Seeing the operation of face detection as a template-matching task suggests many new questions. For example, what does the face template look like and how is it size invariant? Research into automatic face detection may offer testable accounts such as a deformable-luminosity template. This research represents an important development in our understanding of human face detection. Obviously, further research will be required to test and develop the ideas offered here before we can be confident that we understand how we perform the difficult yet automatic task of face detection. It should be noted that the results of our experiments might equally well be found for the detection of other objects. It may be the case, therefore, that understanding face detection allows for a greater understanding of object recognition in general.

References
Bentin S, Allison T, Puce A, Perez E, McCarthy G, 1996 "Electrophysiological studies of face perception in humans" *Journal of Cognitive Neuroscience* **8** 551 – 565
Brown V, Huey D, Findlay J M, 1997 "Face detection in peripheral vision: do faces pop out?" *Perception* **26** 1555 – 1570
Burton A M, Bruce V, Hancock P J B, 1999 "From pixels to people: a model of familiar face recognition" *Cognitive Science* **23** 1 – 31
Collishaw S M, Hole G J, 2000 "Featural and configurational processes in the recognition of faces of different familiarity" *Perception* **29** 893 – 909
Cooper E E, Wojan T J, 2000 "Differences in coding of spatial relations in face identification and basic-level object recognition" *Journal of Experimental Psychology: Learning, Memory and Cognition* **26** 470 – 488
Eimer M, 1998 "Does the face-specific N170 component reflect the activity of a specialized eye processor?" *NeuroReport* **9** 2945 – 2948
Elgavi-Hershler O, Hochstein S, 2002 "Vision at a glance: a high-level pop-out effect for faces" *Perception* **31** Supplement, 20
Farah M J, Wilson K D, Drain M, Tanaka J W, 1998 "What is 'special' about face perception?" *Psychological Review* **105** 482 – 498
Gong S, McKenna S J, Psarrou A, 2000 *Dynamic Vision: From Images to Face Recognition* (London: Imperial College Press)
Han C-C, Liao H-Y M, Yu G-J, Chen L-H, 2000 "Fast face detection via morphology-based pre-processing" *Pattern Recognition* **33** 1701 – 1712
Hay D C, Young A W, 1982 "The human face", in *Normality and Pathology in Cognitive Functions* Ed. A W Ellis (London: Academic Press) pp 173 – 202
Hill H, Watt R, 1996 "Face detection in natural scenes" *Perception* **25** Supplement, 44
Hole J A, George P A, Dunsmore V, 1999 "Evidence for holistic processing of faces viewed as photographic negatives" *Perception* **28** 341 – 359
Johnston A, Hill H, Carman N, 1992 "Recognising faces: effects of lighting direction, inversion, and brightness reversal" *Perception* **21** 365 – 375

Juell P, Marsh R, 1996 "A hierarchical neural network for human face detection" *Pattern Recognition* **29** 781 – 787

Kemp R, McManus C, Pigott T, 1990 "Sensitivity to the displacement of facial features in negative and inverted images" *Perception* **19** 531 – 543

Lee C H, Kim J S, Park K H, 1996 "Automatic human face detection in a complex background using motion and colour information" *Pattern Recognition* **29** 1877 – 1889

Lewis M B, Edmonds A J, 2002 "Localisation and detection of faces in naturalistic scenes" *Perception* **31** Supplement, 19

Lewis M B, Edmonds A J, forthcoming "Searching for faces in naturalistic scenes", submitted to *Visual Cognition*

Lewis M B, Gerhand S, Ellis H D, 2001 "Re-evaluating age-of-acquisition effects: are they simply cumulative-frequency effects?" *Cognition* **78** 189 – 205

Lewis M B, Johnston R A, 1997 "The Thatcher illusion as a test of configural disruption" *Perception* **26** 225 – 227

Loftus G R, Masson M E J, 1994 "Using confidence intervals in within-participants designs" *Psychonomic Bulletin & Review* **1** 476 – 490

McClelland J L, 1979 "On the time relations of mental processes: An examination of systems of processes in cascade" *Psychological Review* **86** 287 – 330

Maurer D, Le Grand R, Mondloch C J, 2002 "The many faces of configural processing" *Trends in Cognitive Science* **6** 255 – 260

Miller J, Ham F van der, Sanders A F, 1995 "Overlapping stage models and reaction time additivity: Effects of the activation equation" *Acta Psychologica* **90** 11 – 28

Nothdurft H-C, 1993 "Faces and facial expressions do not pop out" *Perception* **22** 1287 – 1298

O'Donnell C, Bruce V, 2000 "The Batman effect: Selective enhancement of facial features during familiarisation" *Perception* **29** Supplement, 76

Purcell D G, Stewart A L, 1986 "The face-detection effect" *Bulletin of the Psychonomic Society* **24** 118 – 120

Purcell D G, Stewart A L, 1988 "The face-detection effect: Configuration enhances perception" *Perception & Psychophysics* **43** 355 – 366

Purcell D G, Stewart A L, 1991 "The object-detection effect: Configuration enhances perception" *Perception & Psychophysics* **50** 215 – 224

Ro T, Russell C, Lavie N, 2001 "Changing faces: A detection advantage in the flicker paradigm" *Psychological Reports* **12** 94 – 99

Sagiv N, Bentin S, 2001 "Structural encoding of human and schematic faces: Holistic and part-based processes" *Journal of Cognitive Neuroscience* **13** 937 – 951

Sternberg S, 1969 "The discovery of processing stages: extensions of Donders' method" *Acta Psychologica* **30** 276 – 315

Tanaka J W, Farah M J, 1993 "Parts and wholes in face recognition" *Quarterly Journal of Experimental Psychology A* **46** 225 – 245

Treisman A, Gelade G, 1980 "A feature integration theory of attention" *Cognitive Psychology* **12** 97 – 136

Turk M, Pentland A, 1991 "Eigenfaces for recognition" *Journal of Cognitive Neuroscience* **3** 71 – 86

Valentine T, 1988 "Upside-down faces: A review of the effect of inversion upon face recognition" *British Journal of Psychology* **79** 471 – 491

Valentine T, 1991 "A unified account of the effects of distinctiveness, inversion and race in face recognition" *Quarterly Journal of Experimental Psychology A* **43** 161 – 204

Valentine T, Bruce V, 1986 "The effects of distinctiveness in recognizing and classifying faces" *Perception* **15** 525 – 533

Vuilleumier P, 2000 "Faces call for attention: evidence from patients with visual extinction" *Neuropsychologia* **38** 693 – 700

Wu H, Chen Q, Yachida M, 1999 "Face detection from colour images using a fuzzy pattern matching method" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21** 557 – 563

Yang M-H, Kriegman D, Ahuja N, 2001 "Detecting faces in images: A survey" *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **24** 34 – 58

Yin R K, 1969 "Looking at upside-down faces" *Journal of Experimental Psychology* **81** 141 – 145