Arguments against a 'configural processing' account of familiar face recognition

A. Mike Burton[a,c], Stefan R. Schweinberger[b],

Rob Jenkins[a] and Jürgen M. Kaufmann[b]

[a]University of York, UK

[b]Friedrich Schiller University, Jena, Germany

[c]University of Aberdeen, UK

Correspondence concerning this article should be addressed to A. Mike Burton, Department of Psychology, University of York, York YO10 5DD, UK. E-mail: mike.burton@york.ac.uk or Stefan R. Schweinberger, Friedrich-Schiller University Jena, Department of General Psychology and Cognitive Neuroscience, Am Steiger 3/Haus 1, 07743 Jena, Germany. E-mail: stefan.schweinberger@uni-jena.de,

Abstract

Face recognition is a remarkable human ability, which underlies a great deal of our social behaviour. We can recognize family members, friends and acquaintances over a very large range of conditions, and yet the processes by which we do this remain poorly understood, despite decades of research. Although a detailed understanding remains elusive, face recognition is widely thought to rely on 'configural processing', specifically an analysis of spatial relations between facial features (so-called second-order configurations). In this paper, we challenge this traditional view, raising four problems: (i) configural theories are under-specified; (ii) large configural changes leave recognition unharmed; (iii) recognition is harmed by non-configural changes; (iv) in separate analyses of face-shape and face-texture, identification tends to be dominated by texture. We review evidence from a variety of sources and suggest that failure to acknowledge the impact of familiarity on facial representations may have led to an overgeneralization of the configural account. We argue instead that second-order configural information is remarkably unimportant for familiar face recognition.

Keywords: Cognition, Mental Representation, Perception, Face Recognition, Configural Processing

Introduction

Face recognition is a fundamental human ability, which lies at the heart of our social world. We can recognize the people we know, family, friends and colleagues, apparently without effort and across a huge range of conditions. This ability underlies our everyday interactions, and allows us to tailor our behaviour continually. However, the ease with which we can recognize each other hides the difficulty of the task: how can we recognise a family member over changes in age, emotion and view?; how can we recognize the person in a photograph, a home movie or in the flesh?; how can we recognize the person in poor lighting or an unexpected place? Although face recognition is not perfect over all these circumstances, it is nevertheless remarkably good, and errors are rare by comparison to successes.

Face perception has become a major focus for psychological research, and the topic attracts interest from a wide range of scientists: social, developmental, cognitive and perceptual psychologists as well as neuropsychologists and neuroscientists. It is therefore perhaps surprising that we still know so little about the central question: how do we recognize the people we know?   In this article, we suggest that one reason for slow progress in this field lies in the willingness to recruit a misleading idea. 'Configural processing' is a key concept in face research. The central idea is that faces differ *both* in their constituent features (my mouth is different from your mouth), and also in their spatial layout (the distance between my mouth and nose is different from yours). Perception of the spatial layout underlies 'configural processing', and is held to be critical for recognising familiar faces. We argue below that the concept is problematic. After more than thirty years of use it remains poorly specified, and vague definitions have perhaps contributed to its general appeal, rendering it hard to challenge. However, despite being difficult to pin down, we argue that there is now enough evidence to make the case that facial configurations are surprisingly *unimportant* in recognizing the faces of those people we know.

Before we begin our critique of configural processing, it is important to point out that the study of face perception is much broader than the problem of familiar face recognition. As well as signalling identity, faces provide the viewer with information about expressions, facial speech, gender and focus of attention (via eye gaze). Faces can be judged more or less attractive, and to display particular personality traits. There has been considerable progress in understanding the perception of these signals. For example, the Oxford Handbook of Face Perception (Calder, Rhodes, Johnson, & Haxby, 2011) provides a comprehensive survey of the advances in a great many sub-fields of face perception. Following an influential early analysis of face perception by Bruce and Young (1986) many researchers are clear that different facial judgements can be based on different sources of information; for example, there is no logical reason why the information used to decide that a face is smiling is the same as the information used to decide that it is Bill Clinton. In fact, there are good reasons to propose that these sources of information are to some extent different: we can judge a smile in both familiar and unfamiliar faces, just as we can judge gaze-direction and facial speech.

Because there are very many judgements that can be made on *all* faces, it is important to be clear that we are concerned here only with the recognition of *familiar* faces (i.e. how can I recognize a picture of Bill Clinton?). Because we plan to be critical of configural processing, it is important to spell out that we are only critical of it in this particular context. The concept has been recruited to explain a very wide range of phenomena, including all the various judgements listed above. In this article, we concentrate on the specific claim, commonly made in the literature, that familiar faces are *recognized* using the spatial layout of their features. By recognition we mean the process of assigning an identity to any image of a known person, for example, 'that's Bill Clinton', 'that's my wife' or 'that's the person who works in the coffee bar'. This process is sometimes called 'identification' and sometimes 'individuation', but for our purposes all these terms are equivalent, and we will use recognition throughout.

Configural Processing: history and definition

We do not provide a thorough review of the literature on configural processing here. There are already many good overviews (for example McKone & Yovel, 2009; Piepers & Robbins, 2012; Tanaka & Gordon, 2011). However, it will be helpful to recap the origins of configural processing, and how the concept gained such wide currency. Early work on face processing tended to emphasize a distinction between featural and configural aspects of faces. Although neither of these terms was well defined, appeal was made to the everyday understanding of features (eyes, noses, mouths), and an intuitive notion of their configuration – that is, their spatial layout. Early observations ruled out the possibility that face recognition relies entirely on individual facial features. For example, familiar faces can be recognized from severely blurred images, in which it is difficult to see features clearly (Harmon, 1973). Even when viewers *can* see facial features clearly, early studies showed considerable sensitivity to subtle changes in their arrangement (Haig, 1984).

It is well-established that faces are hard to recognize when presented upside-down, a phenomenon known as the 'inversion effect' (Yin, 1969). Attempts to understand this effect have been closely linked to configural processing. In addition to harming recognition, it turns out that inverting a face reduces viewers' sensitivity to the spatial layout of features (Sergent, 1984). This finding led Diamond and Carey (1986) explicitly to hypothesize that "the large effect of inversion on face recognition results from the fact that faces are individuated in terms of relational distinguishing features" (ibid., p. 108).

To examine configural processing accounts of face recognition, we recruit the careful definitional distinctions made by Maurer et al. (2002), and adopted widely. These authors distinguish three types of configural processing: (i) detection of 'first-order' relations, which define the basic arrangement of a face, that is the fact that face detection relies on a lay-out of features such that two eyes appear above a nose, which lies above a mouth; (ii) holistic processing, which coheres the features into a perceptual gestalt; and (iii) sensitivity to second

order relations, that is the specific spatial arrangement within the face or "perceiving the

distances between features". These three types of processing are tapped by different

perceptual tasks, and Maurer et al. demonstrate that they are behaviorally dissociable (though

all of these are held to be affected by inversion).

Although Maurer et al.'s analysis has been influential, there is still some lack of clarity

in the literature about what precisely is meant by configural processing. Some authors use the

terms 'holistic' and 'configural' interchangeably, and some are unclear about what form of

configural processing is being recruited in an explanation for a particular effect.   Because we

are, in this paper, concerned only with a configural account of familiar face *recognition*, we

will adopt Maurer's terminology, and make it clear that we are addressing only second-order

configural processing here. This is an important specification as, fortunately, some authors

have been clear in framing the hypothesis linking second-order configural processing to

recognition. For example, Richler et al. (2009), write: "Because faces are made from

common features (eyes, nose, mouth, etc.) arranged in the same general configuration, subtle

differences in spatial relations between face features being encoded [are] particularly useful

for successful recognition of a given face." (p. 2856). These 'subtle differences in spatial

relations between features' are precisely those identified by Maurer et al. as second-order

configural properties. Tanaka and Gordon (2011) are likewise clear in their definition: "We

use the term 'configural processing' … to refer to encoding of metric distances between

features (i.e. second-order relational properties)" (p. 178).

Popularity of Configural Processing Explanations

Why is face recognition so often explained in terms of configural processing? One

contributing factor is that there is rather good evidence for the involvement of holistic

processes – the precedence of the whole face over its parts (see below). For example

individual features are best remembered when they are embedded in a complete face, rather

than in isolation (Tanaka & Farah, 1993). Furthermore, when the top and bottom halves of

two different faces are aligned, these tend to fuse perceptually into a novel identity (Young et al., 1987), which in turn hampers the separate processing of individual halves relative to a non-aligned arrangement. This is known as the 'composite effect', and strongly suggests that face recognition recruits the entire percept, rather than extracting local features independent of one another. The fact that viewers tend to see composite faces as a single person implies a role for one type configural processing (holistic), but it does not necessarily imply a role for the involvement of second-order processing.

Another contributory factor may be that evidence is sometimes recruited from face perception tasks other than recognition. For example, there are very many studies examining the effects of configural changes on unfamiliar face image-matching (i.e. judging whether two images are identical), and these normally demonstrate reduced sensitivity following inversion (e.g. Friere, Lee & Symons, 2000; Le Grand, Mondloch, Maurer & Brent, 2001; Rossion, 2008). Furthermore, there is debate about the exact nature of inversion effects, particularly as they affect configural processing (e.g. Goffaux & Dakin, 2010; Sekunova & Barton, 2008). However, all these studies examine a rather constrained task: viewers are asked to report whether two images of *unfamiliar* faces, are *identical or not,* when they are upright or inverted. There seems to be no very compelling reason to assume that the processes involved in that task capture the processes involved in recognizing a friend in the street. Indeed, much evidence attests to dissociations between image recognition and face recognition, and between familiar and unfamiliar face processing. We return to these distinctions below. Whether or not strict processing dissociations exist, familiar face recognition is undeniably a key part of our everyday perceptual experience. One of our central hypotheses is that explanations for different tasks (e.g., identical image matching, or remembering unfamiliar face pictures) may have been over-generalized to account for the phenomenon of everyday recognition, which is both more commonplace, and more difficult to understand.

In addition to its intuitive appeal, a final reason for the popularity of configural accounts is probably the lack of a plausible alternative. If we consider a face to comprise its features (e.g. a particular nose, a particular pair of eyes etc), and their spatial lay-out (e.g. the distances between these features), then we might ask how each of these sources of information contributes to the recognition of that face. Experimentally, one can selectively alter features (e.g. by inserting a new nose) or change their layout (by graphically altering the feature location), and both of these are clearly very important for the percept (for example see research on constructing a forensic likeness, e.g. Frowd et al., 2014). Very early studies taking this approach established that recognition cannot be carried by features alone. (Bradshaw & Wallace, 1971; Matthews 1978; Sergent 1984). These studies showed that the time required for a viewer to judge two faces to be different is not predictable from a simple addition of the component differences. Thus, even in the 1970s and 1980s, there was little enthusiasm for approaches to face recognition which relied primarily on features. More recently, there have been some suggestions that one should consider the role of individual features more carefully (e.g. Cabeza & Kato, 2000; Rakover, 2002).   However, even these modern re-evaluations rely on an integration of featural and configural processing, and present little challenge to the view that configural processes are key to understanding face recognition.

Given the history of face recognition research, one reason to retain a configuration-based hypothesis is the lack of any other useful theory. To avoid any narrative tension, we should point out that we are not going to propose a specific alternative here. Throughout this article we will briefly mention some alternative approaches that warrant further exploration. However, the validity of our critique of the configural approach to recognition does not rely on the reader's accepting any particular alternative.

Four Problems with the Configural Processing Approach

In what follows, we will consider four key problems with configural processing as a mechanism for familiar face recognition: (1) configural theories are under-specified; (2) large configural changes leave recognition unharmed; (3) recognition is harmed by non-configural changes; (4) in separate analyses of face-shape and face-texture, identification tends to be dominated by texture. Problems 2 and 3 are two sides of the same issue, and each of the final three problems arise from well-established empirical evidence. However, we start with a more general problem, under-specificity, which has persisted over many years. In the final section of the paper, we attempt to resolve these problems by suggesting fruitful approaches to future work.

Problem 1: Configural theories are under-specified

If faces are recognized by their 'metric distances between features', then one might expect researchers to operationalize this notion. Exactly which distances are important? In fact, this is never specified. Authors sometimes provide an example, without any commitment to the specific distances used, for example, "nose to mouth distance" and "inter-ocular distance" have been suggested as candidates for key spatial relations (Leder & Bruce, 2000). However, these suggestions are intended only to make a general appeal to an intuitive notion of spatial layout. McKone and Yovel (2009), provide more detailed examples by suggesting landmark points within the face – however, these authors are concerned to explain inversion, rather than identification, and even these points are specifically flagged as being 'theoretical ideas' rather than the actual distances which might be used in face perception.

FIGURE 1 HERE PLEASE

To illustrate the problem Figure 1 shows an attempt to characterize three well-known faces in terms of distance metrics drawn from the literature. The figure shows very clearly that these distances simply do not discriminate between three very different faces. The problem is that distances between features show as much within-person variability (differences between different photos of the same person) as between-person variability

(differences between photos of different people) – meaning these distances are not useful for

discriminating between the individuals. Some of the within-person variability in Figure 1 is

due to changes in facial expression, which affect the location of reference points (e.g. corners

of the mouth). Figure 2 shows a complementary analysis based on standardized interocular

distance – a measure that is invariant across expression. Once again, within-person variability

is as large as between-person variability, presumably due to differences in viewing distance

and properties of the camera lens (e.g. Harper & Latto, 2001). Of course, we understand that

no proponent of configural processing would claim that these particular distances are

necessary or sufficient for discriminating between people. But without some commitment to

a measurement that could be used in this way, it is impossible to evaluate the claim that

metric distances are important in face recognition. In fact, after so many years of use without

specification, one wonders whether the concept could *ever* be operationalized.

FIGURE 2 HERE PLEASE

In the early days of automatic face recognition research, approaches based on the

metric distances between features were tried many times (Kanade, 1973; Kelly, 1970).

However, these consistently failed to produce workable solutions. No group has ever found a

set of facial measures which uniquely identify one person – being similar across instances of

that person, and different from everyone else. Modern-day automatic systems rely much

more heavily on reflectance-based information (see Zhao et al., 2003) and we return to this

approach below. Similarly, an approach to recognition based on measurements between

features ('anthropometry') has been shown to be unworkable in forensic identification

(Kleinberg et al., 2007).   Neither of those fields (computing or forensics) has any

commitment to particular theories of human face perception. They have moved away from

metric distances between features for the entirely practical reason that it does not work. It is,

of course, possible that new research will be published one day that specifies a set of

measurements that are stable enough to be useful in uniquely identifying someone. However,

progress to date has not been promising, and many researchers in face recognition apparently

ignore this fact, recruiting configural processing as a theoretical tool without any

operationalization to use, and no attempt to develop one.

Problem 2: Large configural changes leave recognition unharmed

If recognition relies on 'subtle differences in spatial relations between features', then it

follows that disrupting these spatial relations should harm recognition. However, this is not

borne-out by the evidence. Instead, recognition of familiar faces appears to be remarkably

robust under a range of deformations.

One of the most powerful demonstrations shows that stretching images in the x- or y-

direction (so they are too wide or too tall) has no effect on recognition whatever. Participants

show no reduction in speed or accuracy to make a familiarity judgement when the image is

stretched up to twice its normal vertical height (Hole et al., 2002). This finding is truly

remarkable. Such a transformation deforms almost all 'metric distances between features'.

All angles, ratios of distances, and measures (except those in one dimension) are entirely lost.

A stretch of up to twice the normal height of an image introduces changes to all metric

distance between features which must far exceed the differences in these measures between

people.

It could be argued that a simple linear stretch is not really so profound a distortion of a

facial image. Some non-linear deformations do make face recognition harder (for example,

shearing, in which a rectangular photo is 'slanted' to the left or right; Hole et al., 2002).

Perhaps the visual system is able to undo the effects of linear stretch before the face

recognition system is recruited. However, this is a circular argument. One does not know,

when looking at an image, what its true aspect-ratio should be. In order to re-scale a photo of

Elvis correctly, one would need first to recognize it as Elvis – and to recruit his characteristic

metric distances in order to discount the transformation in these distances. So, the robustness

of recognition across even this rather simple transformation is a considerable challenge to

configural accounts of recognition. This robustness can be found in more fundamental

measures of recognition. An early ERP component sensitive to the repetition of individual

familiar faces, the N250r, is equivalent whether a target face was preceded by a veridical or a

stretched prime face (Bindemann et al., 2008). Somehow, the neural processes involved in

priming a representation of a familiar face are able to discount this severe alteration to the

spatial layout of features.

FIGURE 3 HERE PLEASE

In fact, face recognition is well-established to be robust over much more complex

transformations than linear stretch. Figure 3 (after Harper & Latto, 2001) shows the effect of

simply changing the distance from the camera to the subject. Such a change in perspective is

typically not even noticed by a familiar viewer, but is only apparent when one sees multiple

photos of the same person together. However, it is clear that the spatial layout between

features is changed in rather complex, non-linear ways. Once again, the size of these

transformations produce changes which presumably exceed the differences between people –

a severe problem for a configural account of recognition.

In some recent work, we have been able to test a very simple hypothesis, derived from

a configural processing account of recognition (Sandford & Burton, 2014). Images of faces

are shown to viewers in the wrong aspect ratio (either 'too wide' or 'too tall'). These appear

in a computer window, and the participants' task is simply to resize them, using a mouse to

drag the corner of the window, "until they look right". From a configural theory of

recognition, we hypothesized that this task would be a much easier for familiar than for

unfamiliar faces. If we really differentiate those we know by the 'subtle differences in spatial

relations between face features' then one should be able to adjust these images more

accurately for a known face than for an unfamiliar one, whose spatial arrangement is

unknown.   In fact, across a series of experiments, we found no evidence for this prediction.

In general, viewers were very poor at the task, being satisfied to resize the images rather

inaccurately. In some experiments, participants made *greater* errors for familiar than

unfamiliar faces, while in others there was no difference. Furthermore, the task was shown to

be sensitive to familiarity in other stimuli: when viewers were asked to resize company

trademarks, they were able to do this more accurately for familiar than unfamiliar items.

Thus, far from relying on accurate, detailed knowledge of the metric distances between

features, viewers accept as veridical a very large range in these distances. In summary, it

seems that viewers are very tolerant of distortions to familiar faces – different types of

change, some of them rather large, seem not to harm recognition at all.

Problem 3: Recognition is harmed by non-configural changes

In contrast to the results discussed in the previous section, there are a number of

situations in which recognition is severely impaired by changes which leave facial

configuration unaltered. We consider the effects of photographic negation, line drawings, and

the odd effects of caricature.

It has been known for many years that it is very hard to recognize a face in

photographic negative (Galper, 1970). However, if faces are recognized through the metric

distances between their features, it is not straightforward to explain why this should be so.

Exactly the same information is present in photographic positives and negatives, and viewers

have no difficulty pointing out the eyes, noses and mouths etc. Negation does reduce

viewers' ability to distinguish between two face images with different spatial distances

between features, that is second-order configural properties (Kemp et al. 1990). However, it

is not clear why this should be. Furthermore, the effect is additive to an effect of inversion,

suggesting the two effects have different loci (Kemp et al., 1990). Altogether, it is hard to see

why a system that can extract differences within a plane would be challenged by a

transformation that leaves these distances unchanged. In fact, later explanations of the

photographic negation effect dispense with configural accounts altogether (Bruce & Langton,

1994; Johnston et al., 1992; Kemp et al., 1996). Instead, these accounts place the weight of

the effect in the perception of pigmentation and shape, based on the perceptual assumption

that light shining from above illuminates surfaces in predictable ways. These explanations fit

with an account of recognition that is not tied closely to configuration, but rather relies on

what we will call 'texture' in the following section.

A second transform of interest is the effect of presenting faces as line drawings. For

example, drawings traced from facial photographs, are recognized rather poorly (Davies et al.,

1978; Rhodes et al., 1987). In itself, this is a rather interesting problem: line drawings

preserve spatial layout, so perhaps a configural account of recognition should predict good

recognition. In fact, simple introduction of 'mass', that is black shading introduced by

luminance-thresholding an image, significantly improves performance (Bruce et al., 1992).

This thresholding provides no additional information about the spatial layout of features,

suggesting that the marked improvement in recognition is being driven by non-configural

processing. The boost to performance seems likely to be based on information about the

surface reflectance of the face, rendered at its simplest in this manipulation.

Finally, we consider the effect of spatial caricature – in which the location of points

within a face are exaggerated with respect to a prototype, usually the average of many faces.

The caricature technique produces images in which the distinctive aspects of a face are

enhanced, that is distinctive feature shapes and distinctive distances between these features

are made even more distinctive. Under some circumstances, caricatures of known people are

better recognized than the originals (for an overview, see Rhodes, 1996). This effect is

sometimes taken as evidence for configural processing – though at first sight it is hard to

imagine how this could work. The caricature transformation, by its nature, affects some parts

of the face more than others, introducing a very complex transformation – for example, if

someone has large nose but an average mouth, then the caricature changes the nose

considerably, but leaves the mouth unaltered. Such alterations would seem to change the

metric distance between features quiet considerably, and in rather complex fashion. For the

effect to be consistent with configural processing, it would seem that one would need to define configuration in much more complex ways than 'metric distances between the features'.

In fact, it is interesting to note that spatial caricaturing appears to work best when the image is degraded in some way. The technique was originally developed for line drawings (Brennan, 1985), and caricatured drawings of familiar faces were recognised faster than drawings that preserved the original spatial layout of features (Rhodes, Brennan & Carey, 1987). At that time, photorealistic caricaturing was not available, but the interpretation of poorer recognition performance found for caricatured line drawings (when compared to veridical photographs) may exemplify the strong focus on spatial information for face recognition by researchers at that time: "…these results may simply mean that photographs contain so much more *spatial* information than hand-drawn caricatures that they are more recognizable despite being less distinctive" (Rhodes et al., 1987, p. 475; emphasis ours). In fact, a similar recognition speed advantage for spatial caricatures over undistorted images turned out to be hard to replicate in the case of photorealistic images that contain texture information, unless one makes the recognition task particularly difficult (Benson and Perrett, 1991; Calder et al.,1996; Kaufmann & Schweinberger, 2008; Rhodes et al., 1997). Overall then, evidence from caricaturing provides little support for a strong role of spatial information in the recognition of (undegraded) images of faces.

Problem 4: In separate analyses of face-shape and face-texture, identification tends to be dominated by texture.

In the Introduction, we noted that computational approaches to face recognition typically do not attempt to solve the problem by analysing the spatial relations between features. In fact, many computational approaches separate face 'shape' and 'texture', and a variety of methods have been used to achieve this (Beymer, 1995; Burton et al., 2001; Vetter & Troje, 1995). Figure 4 shows one way in which this operation can be performed. In this

example, shape is defined as a set of anatomical points in an image, and these are marked up

(usually by hand) for all images in the set. A common shape is then defined: often the average

shape of all set members. Using standard face morphing techniques, each face is then

deformed to the same common shape. (Technically, this is achieved by warping the color, or

grey levels, contained in each triangle of the source image, so that it fits the shape of the

corresponding triangle of the target shape, for details see Beale & Keil, 1995.) The resulting

images are often called 'shape-free' faces, because their shape is common – that is one cannot

use simple shape to discriminate between the faces in figure 4b. This procedure allows

separate analysis of shapes (grid points prior to morphing) and 'shape-free faces' in which

shape and feature placement align for all faces in the analysis. The information remaining in

these 'shape-free' images is difficult to name, because it includes texture, color, reflectance,

and information based on the capture device. We use the term 'texture' here as a short-hand

for all these, while acknowledging that other authors use different terms.

FIGURE 4 HERE PLEASE

Separation of face shape and texture is usually performed as a computational

convenience, since it allows faces to be normalized prior to some statistical treatment, such as

principal components analysis (Burton et al., 1999; Calder et al., 2001). However, a few

computational studies have explicitly compared the diagnostic information carried in the

shape and texture components separately, and in these cases information in the texture

components has been found to dominate recognition of identity (Calder et al., 2001; Hancock

et al., 1996; Taschereau-Dumnouchel et al., 2010).

There are also some studies in which the separate contributions of shape and texture

have been tested in the recognition of familiar people, many of these exploiting standard ERP

effects in face perception. Results suggest that shape is important in *learning* a face, but once

learned, plays rather little role in recognition. For example, spatial caricaturing (with texture

unchanged) has a clear ERP effect for unfamiliar faces (more negative occipitotemporal P200

and N250 responses), but no effect on familiar faces (Kaufmann & Schweinberger, 2008).

Recognition performance has also been shown to be better for naturally distinctive faces than

for spatially caricatured faces (Schulz, Kaufmann, Walther, & Schweinberger, 2012),

suggesting that non-spatial distinctive information (i.e., texture) plays a significant role in

recognition. In line with this idea, Itz, Schweinberger, Schultz & Kaufmann (2014) directly

contrasted the effects of spatial caricaturing with texture caricaturing (i.e. exaggerating

luminance and coloration, while leaving shape unchanged). Faces were learned in one of three

versions (veridical, spatially caricatured, or texture-caricatured) and later recognized among

analogous versions of non-learned unfamiliar faces. Recognition performance for learned

faces was best for the texture-caricatured version (when compared to both the spatially

caricatured and veridical versions).

The experiments considered so far in this section look at separation of 'shape' and

'texture' (or 'reflectance') in photographic stimuli. But 2d photographs are generated from

the projection of 3d objects, and some researchers have examined the information carried in

the 3d shape itself (i.e. the pure spatial layout) without any surface texture. For example,

graphically-rendered 3d busts of familiar people are rather poorly recognized by viewers –

and much less-well recognized than photos of faces taken in corresponding poses (Bruce et

al., 1991). More sophisticated technology allows simultaneous capture of 3d shape and

surface reflectance, and it is possible to view each independently (O'Toole et al., 1999;

O'Toole, 2011). Figure 5 shows independent representations of these two face components

(Hill, Bruce & Akamatsu, 1995). Research with these types of images consistently shows that

texture components dominate recognition of identity. For example, in figure 5, the shape of a

particular person is rendered highly accurately, whereas the texture is mapped to a rectangle.

Despite this, people are still better at recognizing the person in the texture than in the shape.

Similarly, when shape and texture are mismatched so that one person's texture is 'wrapped

around' another's shape, viewers' identity judgements seem to rely almost entirely on texture,

with very little effect of 3d shape (see Bruce and Young, 2012). Once again, such

demonstrations suggest that the spatial layout of the 3d features does not provide a strong cue

to recognizing a familiar person.

FIGURE 5 HERE PLEASE

Resolution

We have now provided rather varied evidence suggesting that familiar face recognition

is unlikely to be achieved by configural processing, at least as it is currently understood. We

have raised four problems with the account, the final three of which depend on empirical data.

In this final section, we consider the appeal of configural processing, and attempt to offer a

more positive direction for future research.

'Configuration' is a concept that maps most easily onto a specific instance of a face (a

particular photo say) rather than a generic representation. Given a photo of a face, it is

relatively straightforward to imagine measuring key distances within it. Let us take for

example, the distance between the corner of the nose and the corner of the mouth – simply

because it is one of the distances mentioned by Leder and Bruce (2000). This distance will

clearly change – through changes in expression and speech and, over longer time periods,

changes in health and age. Thus, it seems like a very bad candidate for differentiating

between people – it is so elastic within a person, that this variability would surely exceed that

between people. However, at the core of configural theories of face recognition is the

assumption that, for key distances between features, within-person variability (what is

different about two images of the same person) is negligible when compared to between-

person variability (what is different about two images of different people). This assumption

seems unwarranted. For example, Jenkins et al. (2011) demonstrate much larger within- than

between- person variability on a number of face dimensions, and on the perception of these

(see also Burton et al., 2011; Burton 2013; Jenkins & Burton, 2011). In fact, we hold that it is

a failure to acknowledge within-person facial variability that has misled us into believing that

the problem of face recognition is essentially the problem of distinguishing between two face *images*.

We propose that a full account of face recognition will rely on two fundamental principles: first, the processes involved in perception of familiar and unfamiliar faces are, to some extent, different; and second, in order to understand *familiar* face recognition, it is necessary to understand how faces vary not only *between-person*, but also *within-person*. The two proposals emerge from work showing that unfamiliar face perception relies to a large extent on pictorial representations (Hancock et al., 2000; Megreya & Burton, 2006, 2007). Our perception of an unfamiliar face seems to be tied quite strongly to a particular image of that face. So, when trying to remember someone, or match two images of that person, performance is severely impaired if the two images are not identical (i.e., photos taken at different times, or under different conditions). In contrast, familiar face recognition relies on more robust representations. Our recognition generalizes across a very large range of images – including poor quality images in novel contexts (Burton et al., 1999). This ability to generalize suggests that representations of familiar faces are more abstract than those for unfamiliar faces – they survive surface changes, and seem to rely on properties which can be extracted from a wide range of individual photos.

This distinction between familiar and unfamiliar faces may hold the key to understanding the role of configural processing in face recognition. If one focuses entirely on the problem of telling faces apart, then it is relatively straightforward to imagine how configural processing is an attractive candidate. However, it is only when one sees multiple images of people, all perfectly recognisable, that it becomes clear that configural differences within different images of a person's face can far exceed configural differences between people (see figure 6). We proposed above that second-order configuration is a property of an *image* not of a *face* (figures 1 and 2). Images are unchanging, and it is easy to see how one can make measurements on a particular photo – what is harder to see is how such

measurements generalize across different photos of the same face. Since image-level analysis is prevalent in perception of unfamiliar faces, it may be that configural processing is engaged when the viewer does not know the face, even though it is not critical when the viewer does know the face.

FIGURE 6 HERE PLEASE

We do not intend to develop this theoretical position further here. Our critique of configural processing does not depend on providing an alternative account of face recognition. But neither is it a counsel of despair. We have shown how it is possible to examine different components of the face, without relying on 'metric distances between features'. Indeed, our consideration of texture, in the previous section, is one such approach. Importantly, we note that a rejection of the under-specified configural account does not imply that a holistic approach to face recognition should be abandoned. We simply emphasize that there are many theoretical possibilities, such as those based on image statistics (e.g. Hancock et al., 1996; Turk & Pentland, 1991), in which faces are built of component parts, but these parts all contain information that crosses the whole face. One of the most compelling pieces of evidence for whole face processing is the composite face effect (Young et al., 1987, and see Rossion, 2013 for a recent methodological review), and the arguments presented here are all consistent with a holistic processing account of recognition.

We have been careful to specify that the arguments above refer only to familiar face recognition, and to point out that tasks relying on picture-level analysis are more plausible candidates for a configural approach. However, it is certainly possible that there are other phenomena in face perception which are vulnerable to similar criticisms. For example, while we have been generally supportive of holistic processing accounts throughout this paper, it is true that these too, tend to be poorly specified. We propose that future theoretical proposals need to be much more tightly-specified, in at least two regards. First, any account relying on metric distances should provide an analysis of which specific distances are intended. Without

candidates for an operationalization of a theoretical concept, any such account remains

uncompelling. Second, it is probably time to abandon circularity in these accounts. To take

one example, it is not clear that the configural processing account of inversion is actually an

account of the phenomenon. A statement that configural processes are not recruited when

viewing inverted faces is rather unsatisfying if one cannot say more. What exactly is not

recruited, and why?

One well-studied inversion phenomenon is the Thatcher Illusion (Thompson, 1980;

see the original paper, or Bruce and Young, 2012, for examples) in which inverting the eyes

and mouth of a photo renders it grotesque – a perception that disappears when the entire face

is then inverted. In a recent review, Peter Thompson, inventor of the illusion, commented: 'I

have often been asked why the effect occurs, and have often recited the mantra of configural

coding not being available when faces are turned upside down, but my heart isn't really in the

explanation' (Thompson et al., 2009, p. 932). Thirty years after the effect was first reported,

this reflects a spectacular lack of progress. At the very least, we hope this article will spur

those who disagree with us to provide a more detailed specification of configural processing.

But our true hope is to have persuaded readers that, for familiar face *recognition,* the

configural account is wrong.

References

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*,

    57(3), 217–239.

Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality

    facial caricatures: Implications for the recognition of natural images. *European

    Journal of Cognitive Psychology*, *3*(1), 105–135. doi:10.1080/09541449108406222

Beymer, D. (1995). Vectorizing face images by interleaving shape and texture computations.

    MIT AI Lab memo 1537. Massachusetts Institute of Technology, Cambridge, MA.

Bindemann, M., Burton, A. M., Leuthold, H., & Schweinberger, S. R. (2008). Brain potential

    correlates of face recognition: Geometric distortions and the N250r brain response to

    stimulus repetitions. *Psychophysiology*, *45*(4), 535–544. doi: 10.1111/j.1469-

    8986.2008.00663.x

Bradshaw, J. L., & Wallace, G. (1971). Models for the processing and identification of faces.

    Perception & psychophysics, *9*(5), 443–448.

Brennan, S. E. (1985). Caricature generator: The dynamic exaggeration of faces by computer.

    Leonardo, *18*(3), 170–178.

Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, A. M. (1992). The importance of

    "mass" in line drawings of faces. *Applied Cognitive Psychology*, *6*(7), 619–628.

Bruce, V., Healey, P., Burton, A. M., Doyle, T., Coombes, A., & Linney, A. (1991).

    Recognising facial surfaces. *Perception*, *20*(6), 755–769.

Bruce, V., & Langton, S. R. H. (1994). The use of pigmentation and shading information in

    recognising the sex and identities of faces. *Perception*, *23*, 803–822.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of

    Psychology*, *77*(3), 305–327.

Bruce, V. & Young, A.W. (2012) *Face Perception.* Hove: Psychology Press.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology (2006)*, *66*(8), 1467–1485. doi:10.1080/17470218.2013.800125

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, *102*(4), 943–958. doi:10.1111/j.2044-8295.2011.02039.x

Burton, A. M., Miller, P., Bruce, V., Hancock, P. J. B., & Henderson, Z. (2001). Human and automatic face recognition: a comparison across image formats. *Vision Research*, *41*(24), 3185–3195.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243–248.

Cabeza, R., & Kato, T. (2000). Features are Also Important: Contributions of Featural and Configural Processing to Face Recognition. *Psychological Science*, *11*(5), 429–433. doi:10.1111/1467-9280.00283

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, *41*(9), 1179–1208.

Calder, A., Rhodes, G., Johnson, M., & Haxby, J. (Eds.). (2011). *Oxford handbook of face perception*. Oxford University Press.

Calder, A. J., Young, A. W., Benson, P. J., & Perrett, D. I. (1996). Self priming from distinctive and caricatured faces. *British Journal of Psychology*, *87*, 141–162.

Davies, G., Ellis, H. D., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, *63*(2), 180-187.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107–117.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*(2), 159–170.

Frowd, C. D., Jones, S., Fodarella, C., Skelton, F., Fields, S., Williams, A., … Hancock, P. J. B. (2014). Configural and featural information in facial-composite images. *Science & Justice : Journal of the Forensic Science Society*, *54*(3), 215–27. doi: 10.1016/j.scijus.2013.11.001

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science, 19(4)*, 207-208.

Goffaux, V., & Dakin, S. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, *1*(143). doi: 10.3389/fpsyg.2010.00143

Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception*, *13*(5), 505–512.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330–337.

Hancock, P. J. B., Burton, A. M., & Bruce, V. (1996). Face processing: human perception and principal components analysis. *Memory & Cognition*, *24*(1), 21–40.

Harmon, L.D. (1973). The recognition of faces. *Scientific American, 227 (November),* 71-82.

Harper, B. & Latto, R (2001). Cyclopean vision, size estimation, and presence in orthostereoscopic images. *Presence*, *10*(3), 312–330.

Hill, H., Bruce, V., & Akamatsu, S. (1995). Perceiving the Sex and Race of Faces: The Role of Shape and Colour. *Proceedings of the Royal Society B: Biological Sciences*, *261*(1362), 367–373. doi:10.1098/rspb.1995.0161

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, *31*(10), 1221–1240. doi:10.1068/p3252

Itz, M.L., Schweinberger, S.R., Schulz, C., & Kaufmann, J.M. (2014). Neural correlates of facilitations in face learning by selective caricaturing of facial shape or reflectance. NeuroImage, 102, 736-747. doi:10.1016/j.neuroimage.2014.08.042

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1671–1683. doi:10.1098/rstb.2010.0379

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323. doi:10.1016/j.cognition.2011.08.001

Johnston, A., Hill, H., & Carman, N. (1992). Recognising faces: effects of lighting direction, inversion, and brightness reversal. *Peception*, *21*(3), 365–375.

Kanade, T. 1973. Computer recognition of human faces. Birkhauser, Basel, Switzerland, and Stuttgart, Germany.

Kaufmann, J. M., & Schweinberger, S. R. (2008). Distortions in the brain? ERP effects of caricaturing familiar and unfamiliar faces. *Brain Research*, *1228*, 177–188. doi:10.1016/j.brainres.2008.06.092

Kelly, M. D. 1970. Visual identification of people by computer. Tech. rep. AI-130, Stanford AI Project, Stanford, CA.

Kemp, R. I., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. *Perception*, *19*(4), 531–543.

Kemp, R. I., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues. *Perception*, *25*, 37–52.

Kleinberg, K., Vanezis, P., & Burton, A. M. (2007). Failure of Anthropometry as a Facial Identification Technique Using High-Quality Photographs. *Journal of forensic sciences*, *52*(4), 779–783. doi: 10.1111/j.1556-4029.2007.00458.x

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (1999). Early visual experience and face processing. *Nature*, *410*, 890.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural

information in face recognition. *Quarterly Journal of Experimental Psychology*

Section A, *53*(2), 513–536.

Leder, H., & Carbon, C. C. (2006). Face-specific configural processing of relational

information. *British Journal of Psychology*, *97*(1), 19-29.

doi:10.1348/000712605X54794

Matthews, M. L. (1978). Discrimination of Identikit constructions of faces: Evidence for a

dual processing strategy. *Perception & psychophysics*, *23*(2), 153–161.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural

processing. *Trends in Cognitive Sciences*, *6*(6), 255–260.

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate

perception of features and spacing in faces, and sometimes not? Toward a new theory

of holistic processing. *Psychonomic bulletin & review*, *16*(5), 778–797.

doi:10.3758/PBR.16.5.778

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a

matching task. *Memory & Cognition*, *34*(4), 865–876.

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A

familiarity-based dissociation. *Perception & Psychophysics*, *69*(7), 1175.

O'Toole, A.J. (2011). Cognitive and computational approaches to face recognition. In A.J.

Calder, G. Rhodes, M.H. Johnson & J.V. Haxby (Eds.) *The Oxford Handbook of Face*

Perception, pp:15-30. Oxford: OUP.

O'Toole, A. J., Price, T., Vetter, T., Bartlett, J. C., & Blanz, V. (1999). 3D shape and 2D

surface textures of human faces: the role of. *Image and Vision Computing*, *18*(1), 9–

19.

Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms

"holistic,""configural," and "relational" in the face perception literature. *Frontiers in

Psychology*, *3*, 1–11. doi:10.3389/fpsyg.2012.00559

Rakover, S. S. (2002). Featural vs. configurational information in faces: A conceptual and

empirical analysis. *British Journal of Psychology*, *93*(1), 1–30.

Rhodes, G. (1996). *Superportraits: Caricatures and Recognition*. Hove: Psychology Press.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures:

Implications for mental representations of faces. *Cognitive Psychology*, *19*(4), 473–

497.

Rhodes, G., Byatt, G., Tremewan, T., & Kennedy, A. (1997). Facial distinctiveness and the

power of caricatures. *Perception-London-*, *26*, 207–224.

Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces

happens at a glance. *Vision Research*, *49*(23), 2856–2861.

doi:10.1016/j.visres.2009.08.025

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception.

Acta Psychologica, *128*(2), 274–289. doi:10.1016/j.actpsy.2008.02.003

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of

holistic face perception. *Visual Cognition*, *21*(2), 139–253.

doi:10.1080/13506285.2013.772929

Schulz, C., Kaufmann, J. M., Walther, L., & Schweinberger, S. R. (2012). Effects of

anticaricaturing vs. caricaturing and their neural correlates elucidate a role of shape for

face learning. *Neuropsychologia*, *50*(10), 2426–2434.

doi:10.1016/j.neuropsychologia.2012.06.013

Sekunova, A., & Barton, J. J. S. (2008). The effects of face inversion on the perception of

long-range and local spatial relations in eye and mouth configuration. *Journal of

Experimental Psychology-Human Perception and Performance*, *34*(5), 1129–1135.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, *75*(2), 221–242.

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

Tanaka .J.W. & Gordon, I. (2011). Features, configuration and holistic face processing. In A.J. Calder, G. Rhodes, M.H. Johnson & J.V. Haxby (Eds.) *The Oxford Handbook of Face Perception*, pp:15-30. Oxford: OUP.

Taschereau-Dumouchel, V., Rossion, B., Schyns, P. G., & Gosselin, F. (2010). Interattribute Distances do not Represent the Identity of Real World Faces. *Frontiers in Psychology*, *1*:159. doi:10.3389/fpsyg.2010.00159

Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception*, *9*, 483–484.

Thompson, P., Anstis, S. M., Rhodes, G., & Valentine, T. (2009). Thompson's 1980 paper. *Perception, 38*(6), 921–932. doi:10.1068/ldmk-tho

Turk, M. A., & Pentland, A. (1991). Eigenfaces for recognition. *Journal Of Cognitive Neuroscience*, *3*(1), 71–86.

Vetter, T., & Troje, N. (1995). Separation of texture and two-dimensional shape in images of human faces. In *Mustererkennung 1995* (pp. 118-125). Springer Berlin Heidelberg.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–759.

Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys (CSUR)*, *35*(4), 399–458.

Figure legends

Fig. 1. Metric distances between features for three well known faces in three different photos each. Red lines show distance between the corner of the eye and the edge of the nose (left and right; Leder & Carbon, 2006), distance between the corner of the nose and the corner of the mouth (left and right; Leder & Bruce, 2000), and distance between the nose and the mouth (Leder & Carbon, 2006). Facial landmarks are not normally well defined, leaving ambiguity in their placement. Here we define the corner of the eye as the center of the canthus, the corner of the nose as the lateral extent of the nasal flange, and the corner of the mouth as the lateral extent of the vermillion zone. Nose-to-mouth distance is defined as the vertical length of the philtrum from the procheilon to the nasal septum. Photo size is standardized so that interocular distance (the distance between the center of the left pupil and the center of the right pupil, a-b) is the same for all images. Metric distances between features are expressed as proportions of standardized interocular distance. For all five measures, within-person ranges are large and overlapping, despite constrained pose. Thus, none of the metrics is individuating, even for these very dissimilar faces.

Fig. 2. Metric distances between the eyes for three well known faces in three different photos each. Photo size is standardized so that iris diameter (a-b) is the same for all images. Distance between the lateral edge of the left iris (a) and the lateral edge of the right iris (red line) is expressed as a multiple of standardized iris diameter. Within-person ranges for this metric are large and overlapping, so that the metric is not individuating.

Fig. 3. Top row: Images of the same person taken at difference distances (c. 0.5m to 3m). From Burton (2013). Bottom row: middle and right-most images overlain to standardize right eye and nose between images, respectively. Poor registration shows the extent of the differences between these images, even in apparently stable measures. For example, simply altering the viewing distance leads to large changes in the distances between the eyes, between nose and mouth, and all measures depicted in Figure 1

Fig. 4a. Decomposition of a face image into shape and texture components. From Burton et al., (2005).

Fig. 4b. Two further celebrities whose shape has been morphed to the same common shape template (from Burton et al., 2005). Images depict Susan Sarandon and Sylvester Stallone.

Fig. 5. Facial surface data captured by a 3D scanner. Information is gathered from a camera which rotates around the head. This delivers xyz co-ordinates which can be rendered as a surface (left), and color information which can either be mapped onto that surface, or displayed as a color map (right). Figure from Bruce and Young (2012).

Fig. 6. Ambient photos of UK Prime Minister David Cameron. All images used under Creative Commons or Open Government Licence. Attributions (top row left to right, bottom row left to right): The Department for Culture, Media and Sport [CC-BY-2.0]; Zasitu (Own work) [CC-BY-SA-4.0]; English Foreign and Commonwealth Office [Open Government Licence v1.0]; Richard J. Cole [CC-BY-SA-3.0]; Umakanth Jaffna (Own work) [CC-BY-SA-4.0]; English Foreign and Commonwealth Office [Open Government Licence v1.0]; Willwal.Willwal at en.wikipedia [GFDL]; Russell Watkins/Department for International Development [CC-BY-SA-2.0]; English Foreign and Commonwealth Office [Open Government Licence v1.0]; The Department for Culture, Media and Sport [CC-BY-2.0]; English Foreign and Commonwealth Office [Open Government Licence v1.0]; Xtrememachineuk at en.wikipedia [Public Domain].