# The Role of Familiarity in Face Learning and Face Inversion

Lydia Rebecca Searle

Doctor of Philosophy

University of York

Psychology

April 2024

# Abstract

Our ability to recognise familiar faces is excellent, yet our perception of unfamiliar faces is surprisingly poor. This large difference between familiar and unfamiliar faces is present across many areas of face perception research, such as face learning and face inversion. In the first experimental chapter, the role of top-down cues on face learning was investigated across 4 experiments, by manipulating the presence or absence of top-down information while participants learnt to recognise faces from either low or high variation images. A small benefit of learning faces from highly variable photographs compared to low variation images was found when faces were learnt with top-down information but not without. The largest effects in the first experimental chapter however were not related to the presence or absence of top-down cues as expected, but concerned whether faces were learnt (familiar) or novel (unfamiliar). Top-down cues were helpful but not necessary for face learning, and whether faces were familiar or unfamiliar showed the biggest effects. In the second experimental chapter, the role of stimuli cropping on the face inversion effect was investigated across 6 experiments. It was found that for famous faces, photographs cropped around the face produced large inversion effects, however, whole uncropped images produced much smaller inversion effects. Conversely, for unfamiliar faces, image cropping did not affect the size of the inversion effect. This thesis as a whole suggests that familiar and unfamiliar faces are processed qualitatively differently, and the differences between them can be explained by the nature of their underlying face representations. A new model of face representations is presented to explain the results, bringing together ideas from Valentine's face space model (1991) and Bruce and Young's recognition model (1986).

# Table of Contents

# List of Figures

# Acknowledgements

I firstly want to thank Dr David Pitcher, who has supervised me through two master's degrees and now a PhD. Thank you for your help and guidance over the years. You've picked me up many times when life or academia beat me down. You have always been incredibly supportive and encouraging, making me feel like I could conquer anything. Thank you for giving me 6 weeks off before my wedding, that was one of the greatest blessings in my life. I know you've always wanted what's best for me. Thank you for everything you've done.

Thank you to Prof Mike Burton, for your excellent ideas, your feedback on earlier drafts of these chapters, and for sharing your stats knowledge. It has been a privilege to work with you.

Thank you to my lovely Chris. You have got me through these 3 and a half years. Thank you for being my biggest supporter, for believing in me and cheering me on the whole time. Thank you for hugging and comforting me when I was reduced to tears. Thank you for listening to me externally process ideas at you, to the point that you know the field almost as well as I do. I love you so much. I dedicate this thesis to you.

Thank you to Mum and John, and Dad. You all helped me financially through my undergrad which was the steppingstone to getting here. Thank you for all of the love and support over the years, and for believing I could do anything.

Thank you to Richard Searle, my kind father-in-law, for proofreading chapter 3 for me.

Thank you to Scarlett Syme and Emma Jackson for being great officemates and lovely friends. Covid meant that we didn't get into the office until we were already a chunk of the way through our PhDs, but thank you for your friendship and sharing the experience of being in the same boat.

Thank you to the ESRC for funding my PhD and for a further 3 months of funding due to covid.

Thank you to Prof Beth Jefferies, Dr Philip Quinlan and Dr Tom Hartley for being on my TAP and for your advice.

And the biggest thank you is to Jesus. Thank You for giving me my dream of doing a PhD. Thank You for lovingly guiding me out of academia. I am excited for what we will do together next.

# Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

# Chapter 1 – Introduction

**Are faces special?**

Faces are important social stimuli. Lots of information can be gained by looking at a face, for example one can tell the gender of someone very quickly from their face (Bruce et al., 1993) and their approximate age (Han et al., 2013). Furthermore, judgements about attractiveness and health from faces can influence mate choice (Shackelford & Larsen, 1999). Moreover, you can tell where someone is attending to by following their eye gaze, which is helpful for joint attention (Kawai, 2011) and being warned about potential danger. Faces can aid our understanding of speech by looking at speech facial cues (McGurk & MacDonald, 1976) and they can reveal the emotion a person is feeling internally (Ekman & Oster, 1979). Most interestingly, faces provide the main source of identity information (Bruce & Young, 1986). The question then is are faces special? As they are such a rich source of social information, are there dedicated brain areas and networks for processing faces, and are there perceptual differences between the processing of faces and other classes of stimuli?

The first line of argument about faces being special is that we may have evolved to orient towards and attend to faces. When newborns are presented with a white oval (representing a head) and 3 black circles where the eyes and mouth would be, they attend to and prefer that stimulus compared to if the arrangement of the 3 circles are inverted (Johnson et al., 1991). Furthermore, 12-week-old babies prefer to look at a simple pixelated black and white face drawing with positive contrast in comparison to negative contrast (Mondloch et al., 1999). This shows that young infants have an innate predisposition to attend to face, or face-like, stimuli. Furthermore, perceptual narrowing occurs in infancy, whereby 6 month old infants fixate on a novel face for longer compared to a familiar face, regardless of whether the face is human or monkey, however by 9 months, infants were worse at discriminating between novel and familiar monkey faces and fixated on both for an equal amount of time (Pascalis et al., 2002). This fits with the other-race effect found in adults, whereby faces of a different race are harder to discriminate between compared to own-race faces (Bothwell et al., 1989). Taken together this suggests that there is an innate face processing system present at birth which gets refined by visual experience, e.g., with upright positive-contrast own-race faces.

*Neural evidence*

A key part of this face processing system is the fusiform face area (FFA). In the seminal study by (Kanwisher et al., 1997), participants underwent functional magnetic resonance imaging (fMRI) and passively viewed images of faces and objects. They found a region in ventral occipital cortex which responded more to faces than objects, which they labelled the FFA. Having functionally defined this region of interest (ROI) in the participants, they then showed them faces, scrambled faces, hands, objects and houses, and again found preferential activity to faces in the FFA. In fact, the timecourse in Figure 1.1 shows the signal increasing when faces are presented and decreasing when objects are shown, demonstrating the face sensitivity of the FFA. Furthermore, the increase in FFA activity to faces from baseline compared to the increase to houses from baseline was 6x greater, further demonstrating a strong stimulus preference. The results from this seminal study clearly demonstrate the face sensitivity of the FFA, however, the stronger position that it is face selective is contested (Andrews & Schluppeck, 2004; Gauthier et al., 1999).

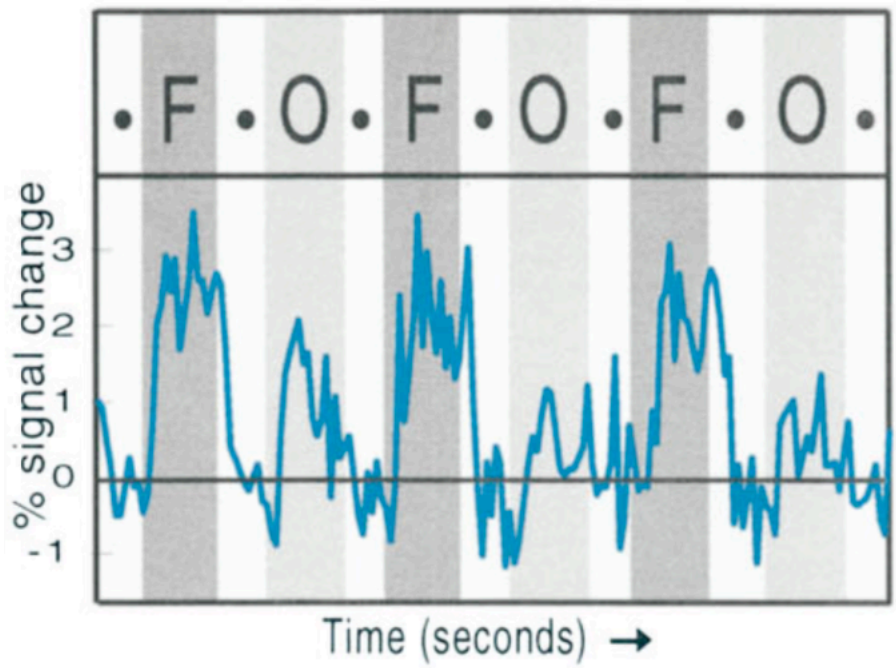



**Figure 1.1.** Timecourse data of the raw percentage signal change in the FFA as faces (F) and objects (O) are passively viewed. Figure taken from Kanwisher et al. (1997).

Electroencephalography (EEG) data supports the MRI findings. In a study by Bentin et al. (1996) scalp electrodes were placed on participants' heads and they viewed multiple images of faces, scrambled faces, cars, scrambled cars and butterflies. They were asked to count how many times a specific category exemplar was shown, e.g. how many faces are there? They discovered the N170, a negative event-related potential (ERP), over the lateral posterior scalp which was sensitive to intact faces. A slightly larger response was found in the right hemisphere compared to the left, consistent with MRI findings that the right FFA is activated more by faces than the left (Kanwisher et al., 1997). The FFA is thought to be the primary neural source of the N170 (Gao et al., 2019). In a follow-up experiment, Bentin et al. (1996) showed participants human faces, animal faces, human hands, cars and furniture and directed participants' attention away from faces by asking them to count the number of cars presented. Still they found the N170 in response to faces, ruling out that it was an effect of attention. Furthermore, the comparisons ruled out that it was face-like stimuli which evoked the N170 as animal faces did not evoke it, nor was it caused by human body parts in general as human hands failed to elicit the response. The N170 is a robust marker of face processing which has been replicated many times since its discovery by Bentin et al. (1996).

Furthermore, neuropsychological evidence also points towards face specific brain processing. Acquired prosopagnosia refers to an impairment in identifying faces usually caused by brain damage, first coined by Bodamer (1947). Prosopagnosia is a very specific deficit of recognising known individuals from their faces. In-depth case studies of patients with prosopagnosia provide insight into face processing. For example, in a study by Monti et al. (2019), patient ST was studied after she suffered a haemorrhage in her right occipital cortex leaving her with acquired prosopagnosia. She did not demonstrate general impairments, for example, she performed within normal ranges on a variety of tests measuring spatial cognition, abstract reasoning, and short- and long-term memory. Furthermore, patient ST did not have an issue with recognition of facial expressions, consistent with findings that prosopagnosics do not struggle to perceive faces or indeed to recognise emotions from faces. While patient ST appears to be quite a pure case of prosopagnosia, it should be noted that some cases of prosopagnosia also present with other face-processing deficits, such as impaired facial expression recognition (Humphreys et al., 2007). However, patient ST was unable to recognise faces familiar to her, for example her husband or son, but she was able to recognise them from other cues, such as eye colour, earrings and beards. This is consistent with the specific impairment of prosopagnosics, that individuals can still be recognised by non-facial cues, it is specifically that they cannot recognise someone by looking at their face (Busigny et al., 2013). Moreover, prosopagnosics are not just impaired in recognising familiar faces but also learning to recognise faces. ST performed significantly worse than controls on a face learning task using the Cambridge Face Matching Task (CFMT). Other case studies further demonstrate the specificity of prosopagnosia. For example, patient WJ could individuate his sheep but not human faces (McNeil & Warrington, 1993). In-depth studies of prosopagnosic patients demonstrate the clear specificity of the face processing impairment, specifically in recognising familiar faces and learning new faces.

We are such experts at perceiving faces and the brain is set up in such a way to process faces, that we see faces where there are not faces to be seen. Face pareidolia is a visual illusion where a face is perceived where there is not one, for example the famous ‘Jesus in toast’ illusion. In a study by Liu et al. (2014) participants viewed faces faintly presented on a noise background, letters faintly presented on a noise background or a pure noise background. While undergoing fMRI, participants classified whether they saw a face, a letter or nothing. In the pure noise conditions they found that participants report seeing a face 34% of the time. Furthermore, they found greater FFA activity when the participants reported seeing a face in the pure noise condition compared to seeing a letter or perceiving nothing. Moreover, there was a significant correlation between how face-like the perceived face was the participants reported seeing and the increase in rFFA activity when participants reported seeing a face in the pure noise condition, demonstrating that the more the participant experienced face pareidolia, the greater the FFA activation. These findings demonstrate that the visual system does indeed have face specific areas, and that we are so tuned to processing faces we will see them even when they are not there.

*Behavioural evidence*

An alternative explanation to the ‘faces are special’ viewpoint, as laid out in the previous pages, is that there are not face specific processing mechanisms but rather expertise specific mechanisms (Diamond & Carey, 1986). In day-to-day life one recognises faces at the individual level, e.g. that’s my friend Emily, however, objects are usually recognised at the category level, e.g. that is a dog, rather than the individual level, e.g. that is Sammy the dog. It has been proposed that the advantage seen with faces over objects is due to our expertise at discriminating faces at the individual level and our lack of expertise at recognising objects at the individual level.

A key study examining this expertise hypothesis is by Gauthier and Tarr (1997) where participants became experts at discriminating objects at the individual level. Participants learnt to identify ‘Greebles’, 3D objects with 4 protruding parts which had slightly different sizes, angles and positions. The Greebles varied along 2 dimensions determining that they could be from different families and of different genders. They were also given individual names. Participants were either trained with these Greebles to be novices (they learnt the names of the protruding parts), or they were trained to be experts (they could identify the family, gender and name of the Greebles, after approximately 7-10 hours of training). At test, in a forced-choice recognition task, participants were shown pairs of Greebles in their studied configuration, in an altered configuration or they were shown isolated Greeble parts. For experts and novices, Greeble parts were recognised more accurately in the context of the Greeble compared to in isolation, just as face parts are recognised better in the context of a face compared to in isolation (Tanaka & Farah, 1993). Furthermore, experts were significantly faster at recognising intact Greebles compared to when the configuration was slightly altered, demonstrating that they had become sensitive to small configural changes, previously thought to be face specific (Tanaka & Sengco, 1997). Some of these results suggest that expertise at within-category object discrimination can produce similar expertise as face recognition. However, whilst they found that experts could recognise intact Greebles faster than configuration altered Greebles, they found no differences in accuracy. Furthermore, they found no cost of inversion in any condition, whereas face inversion produces a large cost to familiar face recognition (Farah, Tanaka, et al., 1995). Such research provides some evidence for the expertise account.

**Face Neural Network**

We have seen that the FFA is a face-sensitive brain region, however, there are multiple areas of the brain involved in face processing forming a network. There are different pathways in the brain for different aspects of face processing. In the influential model put forward by Haxby et al. (2000) they define brain areas in a core and extended face network forming multiple pathways, as shown in Figure 1.2. The core network consists of the inferior occipital gyri, where the OFA is located, for perception of facial features. This area projects to the lateral fusiform gyrus, where the FFA is located, for identity perception. It also projects via another pathway to the STS for emotion recognition and eye gaze perception. The FFA and STS then project to other areas of the brain in other networks for further processing in the extended system.

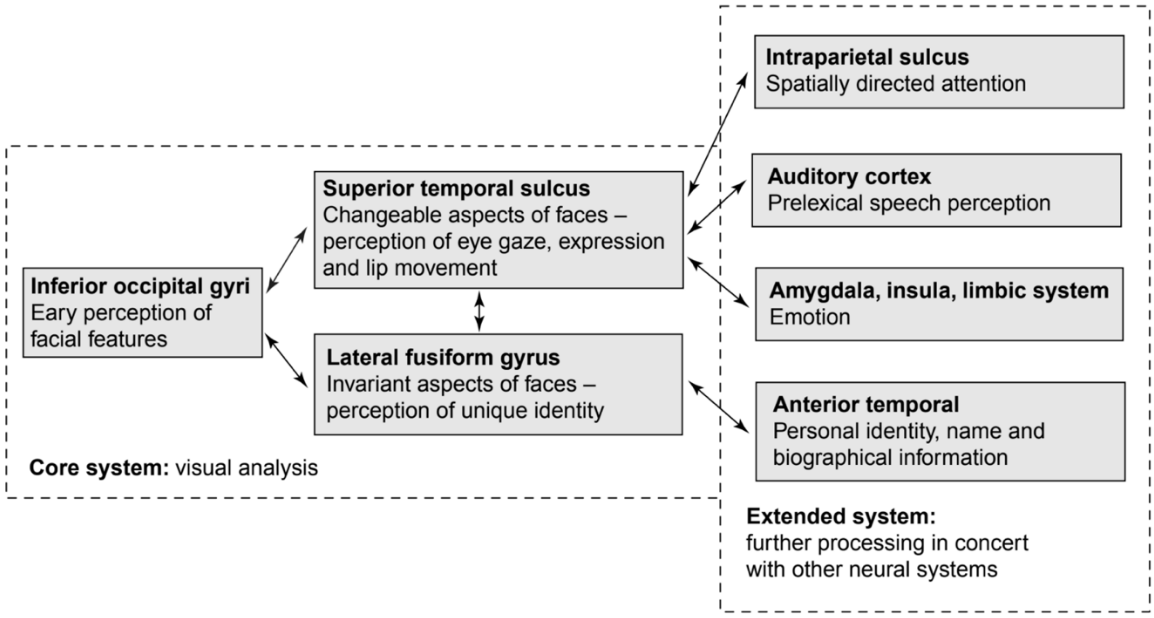



**Figure 1.2.** A model of the distributed human neural system for face perception, taken from Haxby et al. (2000).

The occipital face area (OFA) is another face-selective region in the brain (Gauthier et al., 1999, 2000; Puce et al., 1996). In a study by Pitcher et al. (2007), participants underwent a delayed match-to-sample task. They viewed 2 images of faces where either the spacing of the face parts had changed or the face features had changed, or 2 houses where the spacing of the house parts (e.g. windows, doors) had changed or the house ‘features’ had changed (e.g. a different front door). Participants received repetitive transcranial magnetic stimulation (rTMS) over left or right OFA or at vertex at the onset of the first face. They found accuracy was impaired when TMS was delivered over rOFA when the face parts were changed, and there were no differences in any of the house conditions. This suggests that rOFA is a face-selective area which processes face parts and may be involved earlier in face processing than the FFA in a feedforward hierarchical model for identity recognition.

Furthermore, the posterior superior temporal sulcus (pSTS) is another brain region which responds selectively to faces (Perrett et al., 1992). In a TMS study by Sliwinska and Pitcher (2018) participants completed a delayed match-to-sample task. Participants viewed a face expressing an emotion, then viewed a different face expressing either the same or a different emotion whilst receiving TMS. Participants either received TMS over right pSTS (rpSTS), left pSTS (lpSTS), or vertex, which produced the lowest, middle and highest accuracy scores respectively. Participants also completed a control object matching task and received TMS at the same sites, which produced no differences in accuracy. This study demonstrates the rpSTS is involved in the perception of facial expressions and whilst this is lateralised to a degree, the lpSTS is also involved.

The face network dedicated to identity perception from faces, is modulated by familiarity. In a study by Weise et al. (2019) participants viewed images of personally highly familiar faces and unfamiliar faces while EEG was used to record ERPs. They found a more negative amplitude from 400-600ms when participants viewed highly familiar faces compared to unfamiliar faces, which they dubbed the sustained familiarity effect (SFE). The brain regions dedicated to the perception of faces, and furthermore, the networks specifically involved in identity recognition are modulated by whether one is viewing a familiar or unfamiliar face.

The literature reviewed so far provides a convincing argument that there are indeed brain areas which are sensitive to faces. However, does that mean that faces are special? It depends on the definition of ‘special’. As Valentine (1988) lays out in his review, the question can be split into two sub-questions: are there unique areas of the brain which process faces and is there a unique kind of processing applied to faces specifically? It seems there are areas of the brain which are face sensitive. However, there are also areas of the brain which are object sensitive, body sensitive (Pitcher et al., 2009) and scene sensitive (Pitcher et al., 2019). Faces could then be considered special in that they are processed in distinctly different areas of the brain from other stimuli, however, that is also true of other kinds of stimuli, such as bodies, objects and scenes. Moreover, there is still debate around whether face areas such as the FFA are face sensitive or selective (Andrews & Schluppeck, 2004; Gauthier et al., 1999; Kanwisher et al., 1997).

The other question around whether faces are special or not asks if they are processed in a unique way. Valentine (1988) concludes in his review that they are not. He says that the arguments for faces being processed uniquely comes from the face inversion effect, and the idea that upright and inverted faces are processed qualitatively differently. However, in his review he found that Yin’s (1969) finding of the face inversion effect was not reliably replicated and he found a lack of evidence for upright and inverted faces being processed qualitatively differently. Furthermore, a recent paper has found that faces and other stimuli may be processed perceptually distinctly but conceptually similarly. In an ERP study by Wiese et al. (2023), participants passively viewed personally highly familiar and unfamiliar faces, animals, indoor scenes and objects. In an early 200-400ms time window, they found evidence for distinct processing of the different stimuli types; for example in Experiment 2 they found a significant effect of familiarity in right occipito-temporal regions for faces, but the significant effect of familiarity for scenes was later and more posterior. However, in a later 400-600ms window, they found that all conditions produced similar results, with overlapping regions in right occipito-temporal regions and similar time courses.

The authors suggest that their results reflect early distinct perceptual processing of the different stimulus categories, but that the different stimulus categories to a degree share later conceptual processing. This therefore suggests that faces are perhaps not processed in a unique way. This study fits with Haxby et al.'s (2000) distributed model of face processing, as there are core face-sensitive areas which process faces, and then later processing happens in non-face-specific areas such as anterior temporal cortex. In Wiese et al.'s (2023) study, they found familiarity effects in each stimulus category. This might suggest that rather than faces being special, familiar stimuli are special, in that they are processed in a unique way compared to unfamiliar stimuli.

**Differences in familiar and unfamiliar face processing**

The field of face recognition has progressed slowly (Burton, 2013). At one time, familiar and unfamiliar faces were grouped together into one class of stimuli, much like objects or scenes. This was supported by research which found a module in the brain, the FFA, responded more to faces compared to other classes of stimuli such as houses, objects or body parts (Kanwisher et al., 1997). Alongside early research not differentiating between familiar and unfamiliar faces, methodological issues also slowed the progress of research. Some experimental procedures involved using the same images during learning and at test, confounding image recognition and face recognition, a different and much easier task (Burton, 2013).

Despite this, it has long been established that there are large differences between familiar and unfamiliar face recognition. Bruce (1982) used an old-new recognition task with personally familiar and unfamiliar faces. Participants studied the face stimuli in a learning phase, then at test were shown novel images of the identities which were either shown in the same view as at presentation, or a different view and expression. They found that participants were more accurate and faster at recognising familiar faces compared to unfamiliar faces. They also found that familiar face recognition accuracy was not affected by the change in presentation, however, unfamiliar faces were recognised less accurately when the presentation changed. Furthermore, Bruce and Young (1986) put forward an influential model of face recognition which distinguishes between different processing for the recognition of familiar and unfamiliar faces. When a familiar face is seen, its encoded structural representation is matched to stored structural codes in a recognition unit (visual descriptions of the face), which then activates the relevant person identity node (containing identity-specific semantic information). However, unfamiliar faces do not have stored structural codes as they have not been encountered before, so they cannot be recognised in the same way. Instead, they are temporarily stored and recognised, by structural encoding and directed visual processing, which may involve actively studying a face in order to remember it.

Research has moved to further examining the differences between familiar and unfamiliar face perception, and the use of matching tasks has revealed clear and reliable differences between the two classes of stimuli. The most notable difference is that familiar faces are very accurately matched and unfamiliar faces are not. In face-matching tasks, participants decide whether a pair of photographs depict the same or different individuals. The photos are either 2 photographs of the same person, a match trial, or photographs of 2 different people who look similar, a mismatch trial. With no time limit for viewing the faces, one might expect near-perfect performance (ceiling effects) for both familiar and unfamiliar faces. However, this is not the case. Participants presented with a target photo and an array of 10 unfamiliar faces with unlimited time to examine them correctly identified the target just 70% of the time (Bruce et al., 1999). Even when task demands were reduced to an array of 2 unfamiliar faces, equally low accuracy rates were found (Henderson et al., 2001). However, matching familiar faces produces a very different pattern of results. For example, Bruce et al. (2001) showed participants either 3s videos, three 1s video stills or one 3s video still of identities which participants had to match to an image of either the same or a different identity. Across all conditions, participants were much more accurate at matching familiar identities compared to unfamiliar identities, on average correctly matching familiar targets 92% of the time compared to 74% for unfamiliar identities. Consider Figure 1.3 and determine if the photo pairs depict the same or different people. Matching familiar faces is trivially easy, but matching unfamiliar faces is noticeably more difficult.

a

b

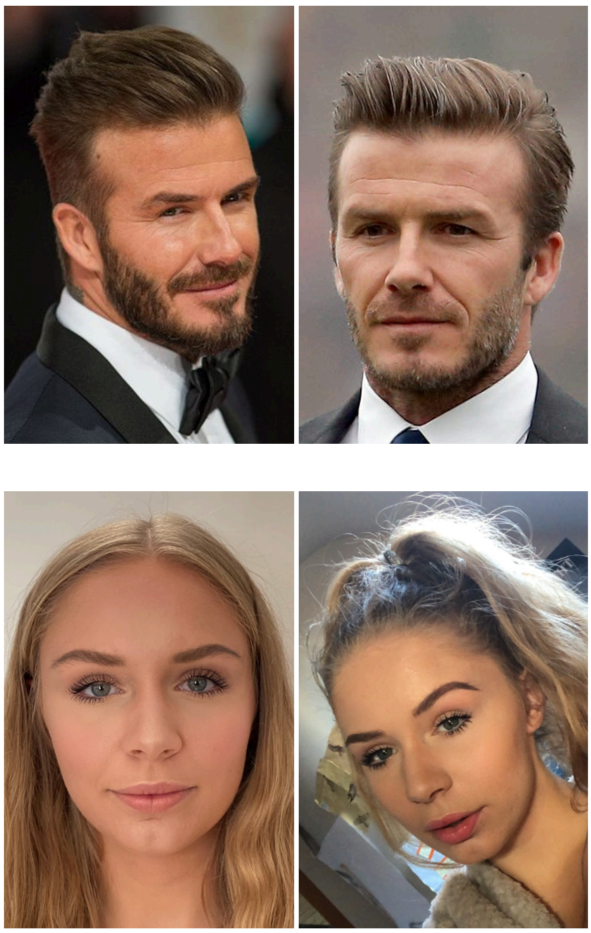


**Figure 1.3.** For each pair, decide whether the photos depict the same individual or not before reading the rest of this caption. a) A familiar face match, b) an unfamiliar face match. Note the speed and ease with which you were able to tell that pair a depicted the same individual, David Beckham. In contrast, note the longer amount of time and difficulty to decide whether pair *b* showed the same person or not, and your confidence with each pair.

Superior processing of familiar faces is also reflected in reaction times (RTs). Matching famous familiar faces produces faster reaction times compared to familiarised faces, which produces faster reaction times compared to unfamiliar faces (Clutterbuck & Johnston, 2004). Ritchie and Burton (2017) also found faster responses when recognising familiar faces compared to recognising unfamiliar identities. This pattern of longer processing times for unfamiliar faces is mirrored in people's confidence judgements. In a study by Burton et al. (1999), participants viewed CCTV video clips of lecturers entering a building, who they were familiar or unfamiliar with. They were then shown high quality images of the lecturers from the video clips and distractors, and rated how confident they were they had seen them. Participants unfamiliar with the targets rated that they were not certain whether they had seen the faces before or not, for both lecturers and distractors. These slower reaction times and greater uncertainty with unfamiliar faces is in stark contrast to fast effortless familiar face recognition.

Moreover, familiar and unfamiliar face recognition rely more heavily on different parts of the face. In the first study of its kind, Ellis et al. (1979) showed participants photos of faces which were either a whole face, the outer features (e.g. hair line, face shape) or the inner features (e.g. eyebrows, eyes, nose, mouth). For famous faces, participants viewed the images and indicated if they recognised them or not. For unfamiliar faces, they performed an old-new recognition task, where unfamiliar faces were shown during a learning phase, and mixed in with distractor items at test. Participants were able to identify unfamiliar faces equally well when given internal or external features, however, when viewing familiar faces, an advantage for recognition was seen with internal features. More recent research has successfully replicated this; Bonner et al. (2003) and Clutterbuck and Johnston (2005) showed internal features of familiar faces were matched faster and more accurately than internal features of unfamiliar faces, and no differences are found between familiar and unfamiliar faces for external features, thus demonstrating another difference in the processing of familiar and unfamiliar faces.

Furthermore, viewers familiar with a face can tell multiple images of the identity together, however viewers unfamiliar with the identity cannot. In the first study of its kind, a card sorting task was developed to test the effect of within-person variation on familiar and unfamiliar face perception (Jenkins et al., 2011). Twenty photographs of 2 Dutch celebrities, which varied in terms of lighting, viewpoint, expression etc., were printed into individual photo cards. These were ambient images as they preserved some of the natural within-person variation. Participants, who were either British and therefore unfamiliar with the faces or Dutch and familiar with the faces, sorted the photos into identity piles, 1 identity per pile. British unfamiliar viewers made a mode of 9 piles, perceiving there to be 9 identities, and very rarely put the 2 different identities into the same pile, showing they had no problem telling the celebrities apart, but they did have problems telling the identities together. In stark contrast, the Dutch familiar viewers made a mode of 2 piles, correctly perceiving there to be 2 identities. This study demonstrates visual processing of unfamiliar faces cannot tolerate within-person variation, as unfamiliar viewers split 1 identity into multiple piles, unable to tell people together. However, familiar face recognition can tolerate within-person variation as familiar viewers were able to correctly sort the varying photos of 1 identity into 1 pile.

Furthermore, in a study by Kramer et al. (2017), in experiment 1 and 2, participants were given 40 cards to sort into identity piles, 1 identity in each pile. The card sets contained 20 photographs of 2 identities, which half the participants were familiar with, and the other half were unfamiliar. Familiar and unfamiliar participants were given 1 of 3 photo card sets, either displaying the full face, just the internal features or just the external features. In the first 2 experiments they found that participants familiar with the faces accurately made a very low number of piles with the full face or internal features photos, correctly telling them together, however, with the external features photos they split the photos into many piles, incorrectly separating identities into multiple identity piles. Interestingly, a similar pattern was found with unfamiliar faces. Participants overall performed worse than the familiar participants, demonstrating the clear familiarity differences shown in Jenkins et al. (2011). However, they again made a large number of piles with the external feature photos.

This contrasts with previous work which has found either no difference between internal and external features for unfamiliar faces, or a slight advantage for external features. The opposite was found here, unfamiliar viewers struggled to tell faces together using external features. The same results were replicated in experiment 2, suggesting that these contradictory findings are a result of the different task used (compared to a matching task). In matching tasks where pairs of photographs are presented simultaneously, participants may be attempting to tell the identities apart, and in such a task format may use the external features in a simple feature-matching way. However, in the card sorting task, participants are focussing on telling people *together* and the external features may not be as useful for telling identities together. In the third experiment, participants who were familiar with the identities were told there were 2 identities present in the photo cards, and therefore to sort the photos into 2 piles. The photo cards were either of internal or external face features. They found that participants were more sensitive to the internal face features and were less accurate when sorting the external face features into piles. This suggests that the external features carry less identity-diagnostic information. Overall, it can be seen there are indeed differences between the reliance upon internal and external features for familiar and unfamiliar faces, however, external features may not carry as much identity-diagnostic information as initially thought.

It would be easy to conclude the differences between familiar and unfamiliar face recognition demonstrated so far are due to memory issues. Previous work has used paradigms which rely on memory, where faces are viewed in a learning phase and then different photos of the same faces mixed with distractors are viewed at test (Leveroni et al., 2000; Duchaine & Nakayama, 2006). Also, in day-to-day recognition, when meeting someone new then seeing them later, one relies on the memory of that first face exposure. However, these differences in familiar and unfamiliar face recognition may not be due to differences in *memory* but rather in *perception*. Face-matching tasks remove memory from the problem, allowing for differences in the perception of familiar and unfamiliar faces to be detected. For example, the matching task data reviewed so far demonstrates the large differences between familiar and unfamiliar faces in the absence of memory-based tasks.

**Unfamiliar faces are not faces**

The perception then of unfamiliar faces is poor, in fact they may not be perceived as familiar faces at all. Megreya and Burton (2006) found that participants matched upright familiar faces accurately, however, when they were inverted, accuracy was impaired, a finding known as the inversion effect. When matching unfamiliar faces upright, participants were less accurate than when matching familiar faces upright, and furthermore no significant impairment of inversion was found. In fact there was a positive correlation between upright and inverted unfamiliar face matching, and between inverted familiar face matching and upright unfamiliar face matching. This suggests that the process by which upright familiar faces are recognised is impaired when the faces are inverted, and as this does not impair unfamiliar face recognition as strongly, a different process is at play for the perception of unfamiliar faces.

Unfamiliar faces may be processed qualitatively differently from familiar faces. For example, Megreya and Burton (2006) also showed that upright unfamiliar face matching positively correlated with object matching. Furthermore, unfamiliar face recognition is impaired by changes in facial expression or viewpoint (Bruce, 1982). Taken together, this suggests that unfamiliar face perception uses less sophisticated perceptual processes which can be disrupted by transient changes to the face and may be more akin to object recognition processes. This is further supported by research by Sekuler et al. (2004) which found participants attended to and around the eyes when recognising unfamiliar faces when upright and when inverted, lending further support to the idea that unfamiliar faces employ a simple feature-based process for identification which is intolerant to variation.

Familiar faces however are perceived and processed in an apparently different way. Unlike unfamiliar face perception, familiar face recognition is robust to changes. Familiar faces can still be accurately recognised when images of faces are stretched to twice their original height (Hole et al., 2002) or when the image is blurred (Sinha et al., 2006). Familiar face recognition is also unimpaired by changes in lighting, viewpoint (Hill & Bruce, 1996) or facial expression (Bruce, 1982). This suggests that familiar face recognition does not use the same unsophisticated perceptual processes as that of unfamiliar face perception, but instead more sophisticated variation-tolerant processes underly it.

Unlike unfamiliar faces, familiar faces may be stored as stable face representations. One possibility is that an average is stored to which input images (seeing a face or photograph) are matched. This account would explain the significant advantage of familiar faces to be recognised accurately despite transient changes in the face. Work by Burton et al. (2005) demonstrated that when multiple images of faces are averaged together, image-level characteristics such as lighting, which vary from image to image and are not identity-diagnostic, are averaged away. This leaves only information which is constant across multiple instances to form a stable face representation of the particular person (see Figure 1.4). They then demonstrated that individuals are better recognised from averages rather than an individual photograph, supporting the idea that familiar faces may be stored as average-based stable face representations.

**Figure 1.4.** The rows of smaller photographs show variable images of 2 identities. The larger images on the right show the product when they have been averaged together. Note image-level characteristics which are not identity-specific are averaged out resulting in a stable face representation. Figure taken from Jenkins and Burton (2011).

In summary, familiar and unfamiliar face recognition are apparently quite different. Familiar faces are recognised more accurately, faster, they rely more on internal facial features, and recognition can withstand large changes in variation. The large differences between these 2 stimuli classes may be underpinned by different perceptual processes, with unfamiliar faces perceived and matched using simple feature-matching strategies and familiar faces being matched against stable face representations. The question then is how are faces learnt? All faces are initially unfamiliar to any given observer and can become expertly recognised. The next sections examine variability and its role in face learning.

**Within-Person Variation**

Face recognition has focused on telling people *apart*. This seems a logical approach to take as all faces share the same first order configuration, a pair of eyes above a nose above a mouth (Maurer et al., 2002), and therefore one must be able to tell a particular face apart from another. This is true of other complex stimuli classes, such as picking up the correct suitcase at an airport amongst many other suitcases or finding the right car in a busy carpark. This is particularly import for face recognition, as recognising and telling faces apart is a crucial social skill used every day with important consequences. The concept of 'doppelgängers', two individuals who look extremely similar, highlights the role of telling people apart in face recognition, for example, Figure 1.5 shows famous musician Justin Timberlake and a doppelgänger who makes a living off his strong resemblance. It is clear then that telling people apart is a crucial social skill.

**Figure 1.5.** The identity on the left is famous musician Justin Timberlake, the identity on the right is a doppelgänger. Images taken from Jetss.com (left) and mirrorimages.co (right).

Experiments often reduce face recognition to photograph recognition (Dyer et al., 2005; Heisz & Shedden, 2009; Kriegeskorte et al., 2007; Rakover et al., 2022). Typically in experiments, a single photograph of an individual is presented multiple times throughout a learning phase. This is based on the premise that a photograph accurately captures the likeness of a face; but this is, perhaps surprisingly, untrue. There are four principal factors which vary in the appearance of an individual. Firstly, there are short-term transient changes in faces, for example different facial expressions, changes in face shape caused by speech, and moving the head in the up-down and left-right planes. Secondly, there are longer-term changes to faces, caused by aging, changing health, variations in skin tone / tan, or the application of make-up. Thirdly, the environment causes changes in face appearance, for example differing light levels and lighting directions illuminates the face and casts shadows in changing ways. These first 3 factors in particular are exemplified in Figure 1.6. Lastly, specifically relating to photographs or videos of faces rather than seeing a face in-person, camera characteristics can impact the look of a face, for example megapixel count, depth of field and focal length can change the appearance of photos of the same individual taken moments apart, as seen in Figure 1.7.

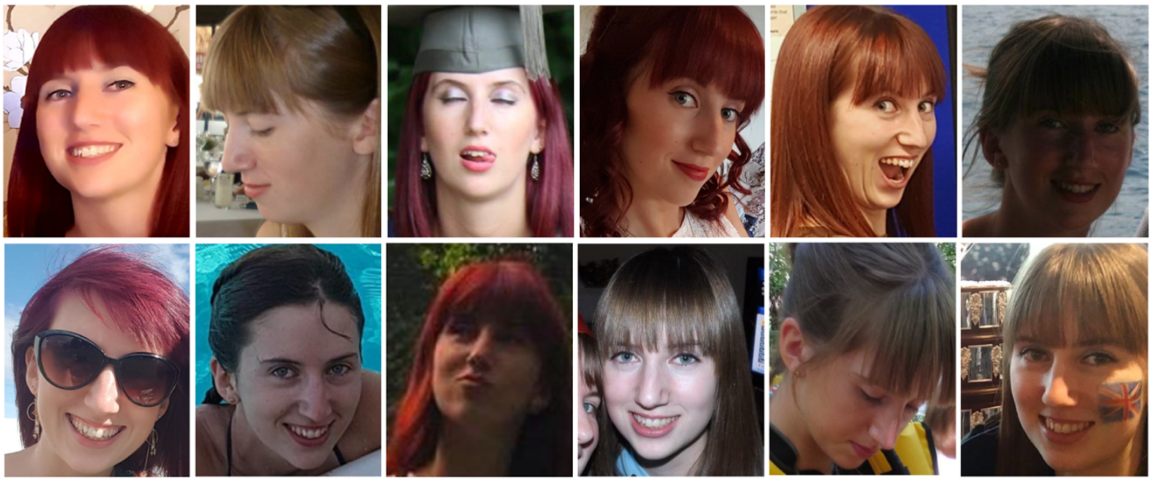


**Figure 1.6.** An array of ambient images depicting 1 individual. The different facial expressions, viewpoints, make-up, ages, hair styles, hair colours, lighting directions, camera characteristics etc. cause the face to look very different.

**Figure 1.7.** All photographs depict the same identity. There is a viewpoint change between the left and right column, and a hair style and camera change between the top and bottom row. All of the images were taken within a few minutes of each other and under the same conditions yet note the differences a change of camera makes to the appearance of a face. Figure taken from Burton (2013).

Research has now quantified within-person variation. In the third experiment of the Jenkins et al. (2011) paper, participants were presented with 12 photographs of 40 different familiar celebrities, and they had to rate how good a likeness each photograph was of the person depicted. If individual faces do not vary, then similar likeness scores should be given to each photo of an individual, however, if within-person variation does exist then a variety of scores would be expected for images of one individual. Mean likeness scores were calculated for each photo, and a high amount of variation in likeness ratings for different photos of the same person were found, as shown in Figure 1.8. In fact, it looks as if there is more variation *within* individuals than *between* individuals. The likeness ratings for each *identity* were averaged, giving a score from 1-7, then these were ranked, giving an identity-rank number from 1-20; the correlations between the mean likeness scores and the identity-ranks for males was 0.936 and females 0.949. Similarly, the likeness ratings for each *image* of an individual were averaged, giving a score from 1-7, then these were also ranked within individuals from 1-20, producing an image-rank number; the correlations between the mean likeness scores and the image-ranks for males was 0.683 and females 0.746. The difference between the rank-identity and rank-image correlations was significant, demonstrating that the variation in the likeness ratings is not solely accounted for by changes in identity, but rather there is variation within-individuals.

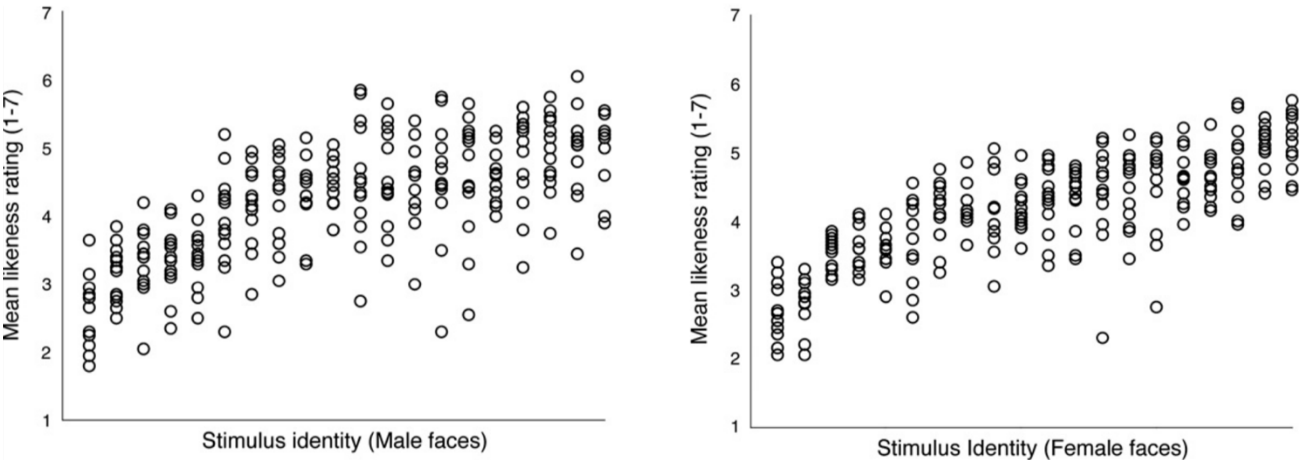



**Figure 1.8.** Mean likeness ratings for individual photos (circles) of different identities (columns), plotted separately for males on the left and females on the right. Taken from Jenkins et al. (2011).

Furthermore, facial attractiveness also reveals within-person variation. In experiment 4 of Jenkins et al. (2011), participants viewed 20 photographs of 20 unfamiliar celebrities and indicated whether they found each face attractive or not. If people do not vary within themselves then one would expect similar ratings for each photograph of an individual. This perhaps would be expected as certain individuals are thought of as attractive and others less so. The number of yes/no responses to each photograph was summed to give an attractiveness score. As in experiment 3, large amounts of variation were seen within individuals, so much so that any pair of photos of different individuals could be chosen to demonstrate that one is more attractive than the other or vice versa. Rank-identity and rank-image correlations were conducted as in experiment 3, and again the variation could not be fully explained by changes in identity, demonstrating that the appearance of an individual's face varies.

We have seen then that there is a vast amount of *within-person variation* observed in a given face (Jenkins et al., 2011). This natural variation is not, however, seen in research, but rather is treated as 'noise' which needs to be controlled away (Burton, 2013; Diamond & Carey, 1986; Kramer et al., 2018; Rakover et al., 2022; Sutherland et al., 2013; Yin, 1969). Experiments typically use highly controlled photographs which are taken on the same day, using the same camera, under the same conditions (Russell & Sinha, 2007). This is in an effort to keep everything constant apart from the variable manipulation. However, this often leads to stimuli being generated for experiments which are grey-scaled with the external features cropped out, and often only 1 photograph is used per identity. While tight experimental control is good scientific practice, in face recognition it has firstly led to unusual stimuli being created which exist only in the laboratory and do not resemble faces seen in day-to-day life, and secondly within-person variation is removed. Jenkins et al. (2011) argue that this disjoint between experimental stimuli and real faces has caused the core issue of face learning and recognition, within-person variation, to be missed.

Since within-person variation has been identified, it has now become a focus of research in its own right. In a study by Ritchie and Burton (2017), participants viewed names paired with multiple images of 10 unfamiliar foreign celebrities, blocked by identity, which were either represented by high or low variability images. The low variability images were stills taken from interview clips, which varied naturally in facial expressions and viewpoint. The high variability images were photographs of the celebrities collected from internet searches, which varied drastically in facial expression, viewpoint, lighting, make-up, hairstyle, camera characteristics etc. After learning the faces and names, participants then completed a matching task, where they viewed pairs of photographs and said if they depicted the same person or different identities. These were either 2 photos of the learned IDs (a match trial) or a learned ID and a foil (a mismatch trial). They found faces learnt from the high variability images were recognised the most accurately and quickly, followed by low variability then novel faces. This study demonstrates that experiencing more variability during face learning leads to better face recognition.

An alternative explanation, however, could be that viewing a range of variable photographs of an identity does not aid face learning. It could be that seeing more photographs increases the likelihood that one will encounter a photograph which looks most similar to the test image, for example they may display the same pose and lighting, and therefore the variable images seen in a learning phase are discarded. Menon et al. (2018) tested whether within-person variation was actually useful for face learning. They showed participants 2 different video clips of an unfamiliar celebrity and manipulated the number of identities the participant thought they were viewing. In the 1 ID condition, they paired both video clips with 1 double-barrel name, e.g. Betty-Sue; in the 2 ID condition, they paired the first video clip with a single-barrel name, e.g. Betty, and the second clip of the same person with a different name, e.g. Sue. They found participants were better able to recognise identities from the 1 ID condition compared to the 2 ID condition. This suggests that the variation seen across the 2 video clips in the 1 ID condition formed 1 face representation, but it was split into 2 face representations in the 2 ID condition, with only 1 clip’s worth of variation in each representation. As there was less variation in the 2 face representations in the 2 ID condition, this led to the lower matching accuracy, showing that within-person is the driving force behind face learning.

The ability to tolerate within-person variation, telling people together, rather than between-person variation, telling people apart, is what causes the large differences between unfamiliar and familiar face recognition accuracy. In a study by White et al. (2022) participants rated how similar 2 images were to each other. In the between-person variation condition, they saw images of face averages of different individuals and the average of those face averages; in the within-person variation condition, participants saw different images of an individual and the average of those images – these were either of unfamiliar or familiar faces. They found no differences between familiar and unfamiliar faces in the similarity ratings for between-person judgements. In contrast, in the within-person condition, they found participants rated individual images of a face and the average as more similar when they were familiar with the identity compared to when viewing unfamiliar faces. As the between-person judgements were not modulated by familiarity, this suggests that the differences between familiar and unfamiliar face recognition are underpinned by within-person variation.

Within-person variation and duration of exposure are 2 potential face learning factors which need to be teased apart. Unsurprisingly, it is known that more time spent viewing a face leads to better subsequent recognition of the face (Memon et al., 2003). We also know this to be true from our personal experience, that we are more likely to recognise someone we have had a long conversation with compared to someone we have only said “hello” to. However, the effects of within-person variation and exposure of duration to a face are sometimes entangled, and they need to be separated. Murphy et al. (2015) directly tested the effects of within-person variation on face learning while holding exposure duration constant. Participants viewed large arrays of variable photographs, containing 6 unique images of 8 identities. Half of the participants saw the same photos in the array in each trial (repeated exemplar learning), and the other half saw new photos in each trial (unique exemplar learning), thereby manipulating the amount of variation the participants saw while holding exposure constant for both groups. In the test phase, they found that the unique exemplar group were significantly more accurate at recognising a novel photograph of the learned identities compared to the repeated exemplar group. This effect was only seen in the match trials and not the mismatch. This study shows that within-person variation does contribute to face learning independently of duration of exposure.

It is clear that within-person variation is critical for face learning, but is it sufficient? In a study by Kramer et al. (2017), participants watched episodes of a TV sitcom in order to learn faces. Watching an episode of a TV program contains natural within-person variation, and is a way faces are learnt in day-to-day life. Participants watched the episodes either upright, inverted, or contrast-reversed. In this design, the exact same amount of within-person variation is seen in all 3 conditions, therefore if within-person variation is sufficient for face learning, subsequent face recognition should be equal across all 3 conditions. At test, participants were shown upright, inverted and contrast-reversed images of the actors from the sitcom and foils, and were asked whether the person depicted appeared in the sitcom they watched. They found when participants viewed the sitcom upright, they accurately recognised upright photos of the actors, but if the photos were inverted or contrast-reversed, recognition was impaired. Furthermore, if the sitcom was watched inverted or contrast-reversed, recognition was impaired whether the face was upright, inverted or contrast-reversed, showing that face learning was disrupted by the unusual viewing conditions. The authors suggest that the perceptual mechanisms underlying face recognition are tuned by our experience, e.g. upright positive-contrast faces, therefore inverted or negative-contrast faces are unable to be processed accurately for identification. Therefore, within-person variation is necessary for face learning, however, it must be within the constraints of what the visual system is accustomed to, and therefore it is not sufficient.

One aspect of within-person variation may be movement. Studies often employ static images in place of live faces, however, when faces are encountered in day to day life they are often moving. Faces show rigid movements such as head nodding and shaking, and non-rigid movements such as those made by speech and facial expressions. This dynamic aspects of faces might also form part of within-person variation and benefit face recognition. This idea was tested by Lander et al. (1999). Participants viewed stimuli which were either degraded by being photographic negatives or inverted positives. The stimuli of the famous identities used were 2.5s long video clips, and 3 stills from the clips shown in total for 2.5s. They found participants were better at naming the faces from moving video clips compared to static images. However, although the viewing time of the videos and images were equated, the video clips contained more static visual information than the 3 stills. Therefore, in a follow-up experiment, participants viewed moving video clips made up of 9 frames, the same frames presented simultaneously side-by-side in the correct order, or the frames presented simultaneously in a random order, thereby equating the amount of within-person variation across conditions while manipulating motion. Again the stimuli were degraded, they were either inverted or thresholded. They found that participants were most accurate at recognising famous faces from moving video clips compared to either of the two static image conditions, demonstrating that motion plays an important role in recognition of faces from degraded stimuli.

Within-person variation is more useful for face learning than face averages. In a study by Dunn et al. (2018) participants viewed an image/s of a target and an array containing the target and distractors and had to locate the target in the array as fast as possible. For the target, participants were either given 1 exemplar photo, a set of 4 photos or an average made up of 19 images. The target image/s and arrays were either of all unfamiliar or all famous familiar faces. They found across all conditions that familiar faces were found more accurately and quickly than unfamiliar faces, and this was not modulated by the target image. For unfamiliar faces, the face average improved accuracy and speed compared to a single image, and a set of 4 photos led to the highest accuracy and fastest reaction times. These findings are supported by similar results which find an advantage for multiple images sets compared to image averages (White et al., 2014). This is particularly surprising as the face average was not made up of 4 photographs like in the 4 photo set condition, but rather of 19 images (or 12 in White et al., 2014) so it contained a more stable representation of the face. Therefore, there is an advantage of viewing natural variation in a face above viewing a stable face average. This may be because the variation seen across multiple photographs is in fact stored in some way, contrary to the averaging hypothesis. Alternatively, it could be because it is the process itself of viewing multiple instances of faces and creating the stored average which leads to better subsequent recognition, rather than just observing a face average.

In previous studies using highly controlled face stimuli, the faces tend to display a neutral expression (Henderson et al., 2001; Clutterbuck & Johnston, 2002). This is also true of official photo identification documents, such as passports and driving licences. However, unfamiliar face matching using such stimuli is known to be highly error-prone in lab experiments (Burton et al., 1999) and in passport border control settings (White et al., 2014). It has been demonstrated that within-person variation is key to face learning, but it may also aid unfamiliar face matching. In a study by Mileva and Burton (2018) participants matched unfamiliar photo pairs. These were either both displaying a neutral expression, a closed-mouth smile or an open-mouth smile. They found there was no difference between the neutral expression and closed-mouth smiles pairs, however, participants were more accurate at matching the open-mouth smile pairs. This suggests that this variation in the appearance of a face, in this case an open-mouth smile, contains idiosyncratic identity-diagnostic information which improves the ability to tell people together, shown in match trials, and to tell people apart, shown in mismatch trials.

We have seen clearly then that faces vary a lot, between individuals but crucially within individuals. This within-person variation is helpful for face learning, and in the next section other factors influencing face learning will be reviewed along with the direction face learning research is heading.

**Face Learning**

Learning to recognise one face does not generalise to the recognition of other faces. In a study by Burton et al. (2016), principal components analysis (PCA) was performed on ambient images of famous actors. This is a statistical technique which derives a multidimensional space where the dimensions are determined by the input faces, and faces can be represented by a relatively low number of components (or eigenfaces). In the first experiment, PCA was performed on 30 images of different famous actors. They found the first 3 components were the same for each identity, coding head rotation in 3D space and potentially lighting. However, from the fourth component onwards, the variation became idiosyncratic, meaning it was specific to each person; people vary in different ways. In a second experiment, they performed separate PCAs on multiple images of different actors, creating a multidimensional space per person. They reconstructed novel images of the identities with either their own PCA components or another's and found that the error was lower when using their own components rather than a same-gendered different actor. This therefore shows that learning to recognise an individual makes you an expert at recognising them, but this learning cannot be generalised to recognising others.

One factor of face learning may be whether the faces learnt are static or moving. In a study by Bonner et al. (2003), participants learnt to recognise faces either by watching short video clips (dynamic) or viewing stills from the video clips (static). They then completed a matching task to test their recognition. On day 1, participants completed the matching task before any learning took place as a baseline, on day 2 and 3 participants completed the learning task then the matching task. One image in the matching task was a full-face image, the other was either of just the internal or external features. They found on day 1 when all faces were unfamiliar, participants were better at matching external features than internal features. Matching accuracy remained similar across days and conditions for external features, however, accuracy increased for internal features over time and particularly so for the dynamic condition. This study replicates the advantage for external features for unfamiliar face matching and the advantage for internal features for familiar face matching. Using this difference, this paper shows that face learning over multiple days leads to better subsequent recognition using internal features, and furthermore that learning faces from moving stimuli leads to better face recognition than static stimuli. However, much more within-person variation was seen in the 30s dynamic clips compared to the 3 image stills used in the static condition. This experiment could be rerun using many photo stills rather than 3 to try capture the same amount of variation as the clips, and the order could be randomised to remove any implied motion, so only static vs moving stimuli is being tested.

Furthermore, the type of motion perceived affects face learning. Lander and Bruce (2003) tested the effects of rigid motion (head moving up, down, left, right) against non-rigid motion (talking and smiling). In one experiment, participants learnt to recognise unfamiliar faces from either rigid motion clips, 5 stills from the rigid motion clips showing the head at different angles, or a single static image. They found at test, participants recognised identities learnt from the rigid motion clips and multiple rigid motion stills equally well, and better than the single static image, suggesting that it was viewing the different perspectives of the face which contributed to face learning, rather than motion. However, in another similar experiment, participants viewed non-rigid motion, again as either short video clips, multiple stills from the clip or a single static image. This time, they found that identities were recognised most accurately if they had been learnt from a non-rigid moving clip, less accurately for the multiple stills condition and least accurately for the single static image, suggesting that seeing faces move in a non-rigid manner is beneficial to face learning.

Furthermore, motion may be beneficial to face recognition if faces have been learnt in motion rather than from a static image. In a study by Lander and Davies (2007), participants learnt to associate names with faces which were either shown as moving clips of a man saying letters and smiling or a freeze frame from the clip. Once participants could name all 12 faces, they completed a recognition test where they viewed either degraded video clips or static images. An advantage for seeing moving faces at test was only found when participants learnt the faces from moving clips. There was no advantage for seeing moving clips at test if the identities were learnt from static images. This suggests that in the moving learning condition, participants were able to extract identity diagnostic information into a face representation which then aided subsequent recognition.

Another potential factor in face learning may be the frequency of exposures, independent of duration of exposure. In a study by Clutterbuck and Johnston (2005) in a face learning task, participants were shown initially unfamiliar faces either 10 times for 2 seconds per presentation, or 5 times for 4 seconds per presentation, keeping the total duration of exposure the same across the 2 conditions. Participants then completed a matching task with whole face and internal/external face pairs, which were either novel identities, the 10 views 2 seconds identities, the 5 views 4 seconds identities or famous familiar faces. There was no difference between the 10 views 2 seconds identities and 5 views 4 seconds groups when matching with external features, but there was a slight increase in accuracy and speed on mismatch trials with internal features for the 10 views 2 seconds identities condition compared to the 5 views 4 seconds. While this was not significant, and no differences were found in the external features matching task, a second experiment comparing novel identities, 10 views 2 seconds identities and famous familiar faces was carried out, and found participants were more accurate and faster at matching internal features of 10 views 2 seconds identities faces compared to novel identities and were more accurate and faster again when matching famous faces. This paper suggested a slight advantage for faces learnt more frequently for shorter periods of time compared to faces viewed less frequently for longer periods of time, however, the findings are weak.

Another potential factor on face learning is eye fixation patterns on the observed face. In a study by Sekiguchi (2011) participants watched video clips of unfamiliar faces while having their eye gaze patterns recorded and gave personality judgements on the observed faces. One week later, participants completed a recognition memory test, in which they saw 20 faces from the learning phase a week prior, or 20 novel faces, and indicated whether they recognised them or not. From the recognition data, the participants were split into a high scoring and low scoring group and their eye gaze patterns were analysed separately. They found across both groups, most fixations were made to the internal features, and this was slightly higher in the high scoring group. The highest number of fixations and the longest durations were on the eyes, then nose, then mouth. The high scoring group fixated significantly longer on the eyes than the low scoring group during face learning, and the low scoring group fixated more times and for longer (but not significantly) on the nose than the high scoring group. This suggests attending to internal features, particularly the eyes, may aid face learning.

A further influence on face learning may be the pairing of faces with semantic information. In a study by Heisz and Shedden (2009) participants learnt to recognise faces while either hearing stories about the identities or hearing unrelated information about volcanoes and rocks. Participants completed this learning task for 5 consecutive days and also completed a passive face viewing task while ERPs were recorded before and after the 5 days of learning. Before the learning phase when all the faces were unfamiliar, they found the N170 repetition effect for unfamiliar faces, where a normal N170 response was observed when different images of the learnt faces were shown, and this response decreased (became less negative) when photos of the same learned identity were repeated. Such an adaptation effect is not found for familiar faces. After the learning phase, they did not find the N170 repetition effect for the learnt identities, but they still found it for novel unfamiliar faces and the learnt identities paired with unrelated stories. Furthermore, 2 months after learning, the majority of participants from the related story group completed a recognition task and all participants recognised all learnt identities. This study shows that when faces are paired with semantically relevant information, they are learnt better than faces paired with unrelated information.

Similarly, processing faces at a semantic level (conceptually) rather than only visually (perceptually) also aids face learning. In a study by Schwartz and Yovel (2019), participants either processed faces perceptually, e.g. by making judgements about the visual appearance of the face such as eyebrow thickness or symmetry, or they processed faces conceptually, e.g. making trait judgements about the faces such as how intelligent or friendly the face looked. Using an old-new recognition task, they found that participants were more accurate and faster at recognising faces which had been processed conceptually during learning compared to perceptually, further demonstrating that pairing faces with semantic information aids face learning.

Context information can also aid face learning. In a card-sorting study by Andrews et al. (2015), participants were given 40 ambient images depicting 2 unfamiliar celebrities. Half of the participants were instructed to sort the photos into identity piles, 1 pile per identity, (free sort), and the other half were told there were 2 identities in the photos and to sort them into 2 piles (2 sort). The free sort group made an average of 6.8 piles, perceiving there to be 6-7 different identities. The 2 sort group created 2 piles as instructed. Half of the participants sorted them perfectly and the remaining participants made only a very small number of ‘intrusion errors’ – an instance where one face appears in a pile where the majority of photos are of the other face. This is perhaps the most interesting finding of the effects on face learning discussed so far, as faces are so variable and only minimal expectation context information was provided, yet participants were able to use this to tolerate within-person variation and tell faces together with similar accuracy to free sorting familiar faces. Context information may play a key role in face learning.

Of the majority of studies reviewed so far, the stimuli and methods used to familiarise faces are not ecologically valid. Standard face learning methods require participants to view a single photograph of an individual repeated multiple times, however, a single photograph does not fully represent an individual's appearance as it does not capture within-person variation (Jenkins et al., 2011). Moreover, studies often use the same image for learning a face and testing recognition, however, this is a large confound in research as image recognition is a different and significantly easier task than recognising the same individual from different images (Burton, 2013). Furthermore, typical face learning methods tend to only display the face for a small number of seconds (e.g. in Clutterbuck and Johnston (2005) study, participants viewed the faces in experiment 1 for a total of just 20 seconds). This is a very small amount of time for a face to make the marked jump from unfamiliar to familiar, and neuroimaging studies suggest this short familiarisation process is not sufficient to achieve the same level of familiarity as famous faces (Leveroni et al., 2000). More ecologically valid methods and realistic stimuli are needed to understand how faces are learnt in the real world.

Recent studies have employed in-person face learning methods. In a study by Sliwinska et al. (2022), participants interacted with to-be-learnt identities in-person for 3 days by having conversations and playing short games. Before the first interaction, participants completed a matching task with the target identities and foils and completed an fMRI scan passively viewing images of the targets and foils. After the in-person face learning, participants completed a matching task and fMRI scan again, using novel images. They found that participants were significantly more accurate at matching images of target identities after the in-person interactions compared to before, however their accuracy for foil identities remained the same. There was no significant difference between matching accuracy of targets and foils before the interactions but there was afterwards. Furthermore, before the interaction no differences were found in neural activity between the targets and foils, however, afterwards target faces activated the FFA and hippocampus more than foil faces. This study employed a highly ecologically valid method as the within-person variation seen in the target faces within each interaction and across each day is akin to meeting someone new and observing such variation in day-to-day life. This method is not only valid as a face learning technique but may also be better than traditional methods as it produced very large improvements in matching accuracy for learnt identities and neural activity akin to famous familiar faces rather than experimentally familiarised faces (Leveroni et al., 2000).

The critique above that face learning studies do not use ecologically valid stimuli and methods is not specific to face learning research. For example, the face inversion effect (FIE) is a well-establish phenomenon in psychology, whereby faces are *disproportionally* affected by inversion, that is the decrease in accuracy when recognising an inverted face compared to upright is significantly larger than the decrease in accuracy when recognising a different stimuli class when inverted compared to upright (Yin, 1969). This effect is long studied, however, the stimuli often used in such studies do not resemble how we see faces in day-to-day life. For example, in the seminal study by Yin (1969), the stimuli are described as being black and white photos of faces cropped under the chin, photos of houses, silhouettes of aeroplanes, stick drawings of men, and sketches of faces and of faceless bodies wearing period costumes. We are told, “The sketches were cropped very severely, so that no hair, ears, or chin lines were present.” (p144). Such stimuli may not reflect our visual experience of faces and other stimuli, yet the use of severely cropped, black and white images of faces are often used in face inversion research (Tanaka & Farah, 1993; Le Grand et al., 2004; Rakover et al., 2022) In the final section of this literature review, I review the face inversion literature, highlighting throughout the need for more ecologically valid methods and realistic stimuli.

**The Face Inversion Effect**

The face inversion effect (FIE) refers to the larger effect of inversion on the recognition of faces compared to other stimuli with a characteristic orientation, or mono-oriented stimuli as referred to by Yin (1969). In the seminal study by Yin (1969) he asked whether the face inversion effect was due to our familiarity with mono-oriented stimuli, and therefore would affect faces and other mono-oriented stimuli equally, or if there is also a special factor relating to faces, which would give rise to larger inversion effects for faces over other stimuli. Participants viewed images of faces, houses, aeroplanes or men in motion during an inspection and test phase. In the first experiment, when the inspection and test images were both inverted, participants produced the most errors when recognising faces from the inspection phase compared to the other stimulus types. This was taken as evidence that although all stimulus types suffered errors when inverted, as faces suffered the most, a special processing mechanism must exist for faces which is particularly inhibited by inversion.

However, in experiment 2, when the inspection images were upright and test images were inverted, the largest inversion effects seen were for men in motion, not faces. As mentioned above, the stimuli used in this seminal paper did not accurately reflect our visual experience with these stimuli classes in real life, and it could be argued that the further away the experimental stimuli were from their real life counterparts, the more errors participants produced. The men in motion produced the highest number of errors and these were stick figures; faces produced the second highest number of errors and these were in black and white and cropped under the chin removing 95% of the person; aeroplanes produced the second least amount of errors and these were silhouettes (I would argued that the faces images were less realistic and more impoverished stimuli than the aeroplanes images, as the planes were not cropped to remove most of the object like with the face stimuli); and houses produced the fewest errors as these were whole photographs, the only issue with those stimuli were that they were in black and white. Despite concerns over the stimuli and one of the experiments producing contradictory results, much research has been conducted on the face inversion effect.

Prosopagnosia provides interesting insights into the face inversion effect. As mentioned above, prosopagnosia refers to the specific impairment of recognising faces. In a study by Farah et al. (1995), prosopagnosic patient LH and a group of neurotypical control participants were tested with upright and inverted faces. LH and controls completed a sequential face matching task with upright and inverted faces. For the neurotypical controls, they found a large decrease in accuracy for inverted face pairs. However, for patient LH, although her overall accuracy was lower, they found an increase in accuracy for inverted faces. The authors suggested that a damaged face-specific processor was automatically engaged by the upright faces, producing poor performance, however, inverted faces did not engage a face processor, and therefore intact object matching strategies could be used to match the inverted faces more accurately.

One explanation given for the face inversion effect is that faces are processed holistically, and it is holistic processing which is inhibited when faces are inverted which gives rise to the effect. The terminology used in the FIE literature is often ill-defined and interchangeable. I will attempt to define key terms in brackets and will use these definitions throughout. Holistic processing, as defined in Maurer et al. (2002) and Rossion (2008) is the parallel processing of multiple interdependent features in a face at once, and the gluing of them into a whole or gestalt. Note that this term is often used interchangeably with configural processing, however, I will later define this term as a separate kind of processing.

Strong evidence and compelling visual illusions demonstrate the holistic processing of faces. The face composite effect was first reported by Young et al. (1987). In their first experiment, they used familiar faces to create face composites, one of the only studies to use familiar faces. The top half of a face belonged to one famous identity and the bottom half belonged to a different famous identity, and they appear to fuse together to form a new facial identity. When the two halves were aligned (a composite), participants were slower at identifying either half of the face, compared to when they were misaligned (non-composites). This suggests that features are not perceived individually, as participants would have performed equally in the aligned and misaligned conditions, but instead this experiment demonstrates that faces are perceived as a whole. In their second experiment, they showed the face composites and non-composites inverted and found that inverted composites were recognised faster compared to upright, demonstrating that inversion, as well as misalignment, inhibits holistic processing. In their third experiment, they used unfamiliar faces to create the face composites and found the same effect, that participants were slower to recognise a target half of the face when aligned with a different face half, however, their effect was smaller than the results of experiment 1 which used familiar faces. This is one of a small number of studies to use familiar and unfamiliar faces, and their results hint at a modulating effect of familiarity. Overall, the face compositive effect provides strong evidence that inversion disrupts holistic processing of faces.

Further evidence that the face inversion effect is due to inhibited holistic processing of faces comes from variations of the face composite task. When two identical top halves of unfamiliar faces are paired with different bottom halves, the tops are perceived as being different, and participants become slower and less accurate at telling the identical top halves apart. The result is found for pairs of faces shown simultaneously (Le Grand et al., 2004) and sequentially (Goffaux & Rossion, 2006). Note however that both of these studies use black and white cropped heads, with either some external features occluded by surgical caps, or all external features cropped out. As can be seen in Figure 1.9, in the top row (A), the top half of the faces look slightly different, as a result of different bottom halves being aligned with them. However, when the faces are inverted in the bottom row (B), the effect is significantly diminished, and it is much easier to tell the top halves of the face are the same.

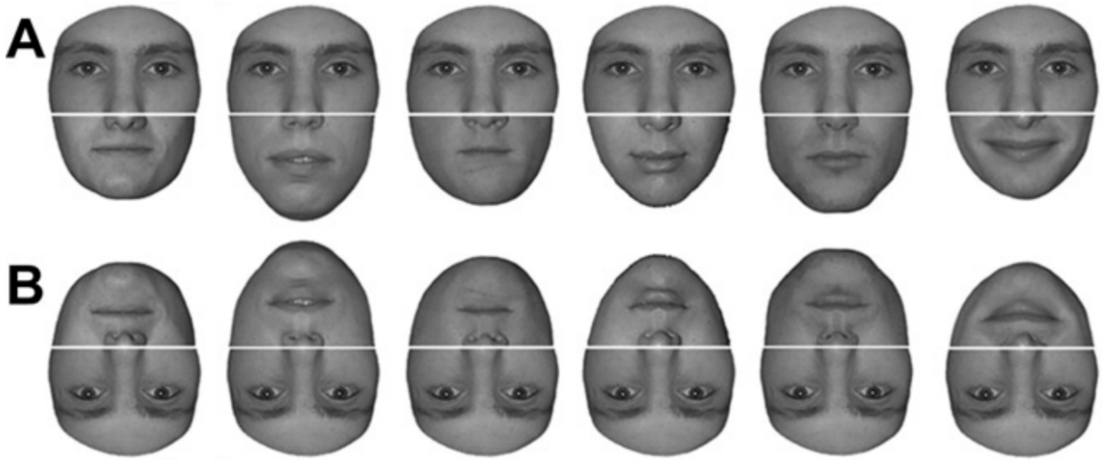



**Figure 1.9.** Figure taken from Rossion (2008) demonstrating the face composite task. In the both the top (A) and bottom (B) rows, the top halves of the faces (the half containing the eyes and eyebrows) are identical, yet look different in the top row and similar in the bottom row.

Holistic processing of faces has also been demonstrated by the part-whole task. In Tanaka and Farah's (1993) first experiment, participants learnt to associate names with faces with the first order configuration (the common layout of faces, eyes above a nose above a mouth) intact or scrambled. At test, participants were tested for their memory of a facial feature by discriminating between pairs of isolated features, pairs of intact faces or pairs of scrambled faces which differed by 1 feature. They found that participants were better at recognising features in the original whole face compared to in isolation; however, for scrambled faces, there was no significant difference between recognising a feature in isolation or in a face. The authors suggest that faces are perceived, stored and recognised holistically, but individual features are not. Therefore, as isolated features and scrambled faces cannot be perceived holistically, but instead are perceived using a feature by feature approach, they do not gain the holistic processing advantage. Furthermore, in their second experiment, they used a similar procedure, but with (intact) upright or inverted faces. They again found upright whole faces were recognised better than upright isolated features, however, no such advantage was found for inverted faces. This again suggests that when faces are inverted, holistic processing is inhibited and instead the features were perceived possibly using a feature-based approach, therefore giving no benefit to the whole face condition. The stimuli used in this experiment were drawings made of dots. Such stimuli do not look like our visual experience with faces and may not be generalisable to real life faces.

In a similar experiment, Tanaka and Sengco (1997) showed participants faces in an ‘original’ configuration, then participants were tested for their memory of a facial feature (e.g. the nose) which was either shown in isolation, in the original configuration or in an altered configuration (eyes further apart / closer together). They found participants were most accurate when the feature was shown in the original configuration, in the middle performance when the feature was shown in the altered configuration and performance was worst when the feature was shown in isolation. The finding that features are best recognised in the original configuration compared to in isolation replicates Tanaka & Farah's (1993) result. Altering the spacing between the eyes worsened performance, however, the feature participants were recognising was the nose. This demonstrates the powerful holistic processing of faces again, that changes elsewhere on the face, irrelevant to the task, affected judgement on the feature of interest. Taken together, these studies reviewed so far suggest that faces are perceived holistically, as a whole, and this processing is inhibited by inversion, leading to the detrimental drops in accuracy behind the face inversion effect.

An alternative account of the face inversion effect is the configural processing hypothesis. This theory states that as faces all share the same first order configuration (eyes above a nose above a mouth), this information only helps to identify a face as a face, but not to discriminate between faces. Therefore, we must use the second order relational information / second order configuration (distances between features) and process faces configurally (*perceiving* the second order configuration) to tell people apart (Diamond & Carey, 1986; Maurer et al., 2002). The theory suggests that objects, however, do not all share a first order configuration and therefore this information can be used for discrimination, second order configural processing is not required. Configural processing then is proposed to be specific to faces and this account hypothesizes that faces will therefore be more affected by inversion than objects. However, a potential issue with this argument in regard to object processing, is that the argument might only work in the context of lab experiments, where faces are identified at the individual level (Chris, Emily etc.) but objects at the basic category level (a water bottle, a building). However, if individual objects were tested (individual water bottles) then first order configuration might well be shared, and differentiation might rely on second order configuration.

The configural account was thought to explain the Thatcher illusion (Bartlett & Searcy, 1993). Thompson (1980) discovered the visual illusion, that when the eyes and mouth are inverted in an upright face, the image looks grotesque, however when this image is inverted (the face is inverted and the eyes and mouth are now upright), the perception of grotesqueness is reduced. However, in a study by Psalta et al. (2014), participants completed a matching task with pairs of faces which were either both normal, both Thatcherized or one normal and one Thatcherized. They found when the pairs were presented upright, participants were very accurate in all conditions, however, when the pairs were presented inverted, participants were less accurate and slower at telling the normal and Thatcherized pair apart. They found the same result when a horizonal slice of the face was presented showing just the eyes or just the mouth. As configural processing of the whole face cannot be carried out in the eyes/mouth only conditions, this study suggests that inversion disrupted local feature processing, rather than inhibiting configural processing.

This configural hypothesis argues that the face inversion effect is qualitative in nature, in that it affects configural processing more than featural processing (perception of local cues, for example the shape / texture of a feature) (Rossion, 2008). A failure of this account, however, is that it fails to explain *why* configural processing would be inhibited more than featural processing. The theory suggests that configural processing is inhibited because we cannot use first order relations to distinguish faces, therefore it says we use second order relations. However, could it not be possible that we use the differences in features (their shape / texture) to distinguish between faces instead / as well? This theory must *assume* that second order configural cues are more diagnostic than featural cues, therefore we rely on configural processing more and do less featural processing. This greater experience of using configural processing on upright faces would lead to greater inhibition for inverted faces, compared to featural processing.

Furthermore, there is evidence that the face inversion effect is quantitative, not qualitative, in nature. Valentine (1988) suggests that if the face inversion effect is qualitative then as a face is rotated away from upright (0°) to inverted (180°), it would produce a non-linear function as a change in processing strategy occurs. However, Valentine and Bruce (1988) tested this idea and found a linear relationship. Participants completed a sequential matching task, where the first image of a face was presented upright, and the second image was either presented upright (0°), or 45°, 90°, 135° or 180° away from upright. For both match (same identities) and mismatch (different identities) they found linear relationships between degrees of rotation away from upright and reaction times. The authors suggested this provides evidence for a quantitative account. However, Rossion (2008) argues that a linear relationship does not rule out a qualitative explanation. He suggests that linear functions, representing for example RTs to detect configural changes or featural changes, could each produce linear functions with differing slopes, and if they interacted that would be evidence for the qualitative explanation.

Despite concerns that the configural account merely describes rather than explains a pattern of data, there is strong evidence for the selective impairment of configural processing in inverted faces. Freire et al. (2000) showed participants faces which varied only in the second order configuration (eyes moved up/down/in/out and/or mouth moved up/down) and asked them to discriminate between pairs either upright or inverted. In experiment 1, they found participants were quite accurate when the faces were presented upright (81%), however, participants were just above chance when viewing the same faces inverted (55%). In their second experiment, they presented participants with faces which had different features (the eyes, nose and mouth varied in each face) and again asked them to discriminate between upright or inverted pairs. This time they found very high and equal accuracy levels for upright and inverted faces (both approx. 90%). The authors suggest that this shows clearly that it is configural processing which is inhibited by inversion and produces the FIE effect, and not featural information. However, there are some issues with this experiment. The stimuli do include the external features and a background, however, they are still in black and white and crop the body out. Furthermore, there are differences in the number of changes made between the configural and featural conditions. In the first experiment examining configural changes, faces have 1 or 2 configural changes made, however, in experiment 2 examining featural changes, faces differ by 3 feature changes. Therefore there are more changes in the featural experiment, making the task easier than the configural changes task. This is demonstrated by the fact that in the upright conditions, participants were approx. 10% more accurate with the featural changes faces.

Further research using a standard set of stimuli (the 'Jane' stimuli) provide more evidence for the configural account of the FIE. Mondloch et al. (2002) edited a model face (Jane) to have either configural changes (the eyes and mouth moved) or featural changes (different eyes and mouth). The first face was presented to participants for 200ms followed by another face until a response was made, and participants had to indicate if they were the same or different. Both faces were either upright or both inverted. They found small inversion effects in the accuracy data for featural changes and larger inversion effects for spacing changes, and no differences were found in the RT data. Although these effects initially seem quite clear, that configural processing is inhibited by inversion while featural processing is spared, it is very difficult, if not impossible, to fully separate these factors in an experiment. For example, in the present study (Mondloch et al., 2002), moving the eyes in the configuration change condition also changes featural information about the nose (making it look longer / bigger). Furthermore, changing the mouth in the featural changes condition also changes configural information, for example the corner of the mouth to the bottom of the nose (see Figure 1.10).

a

b

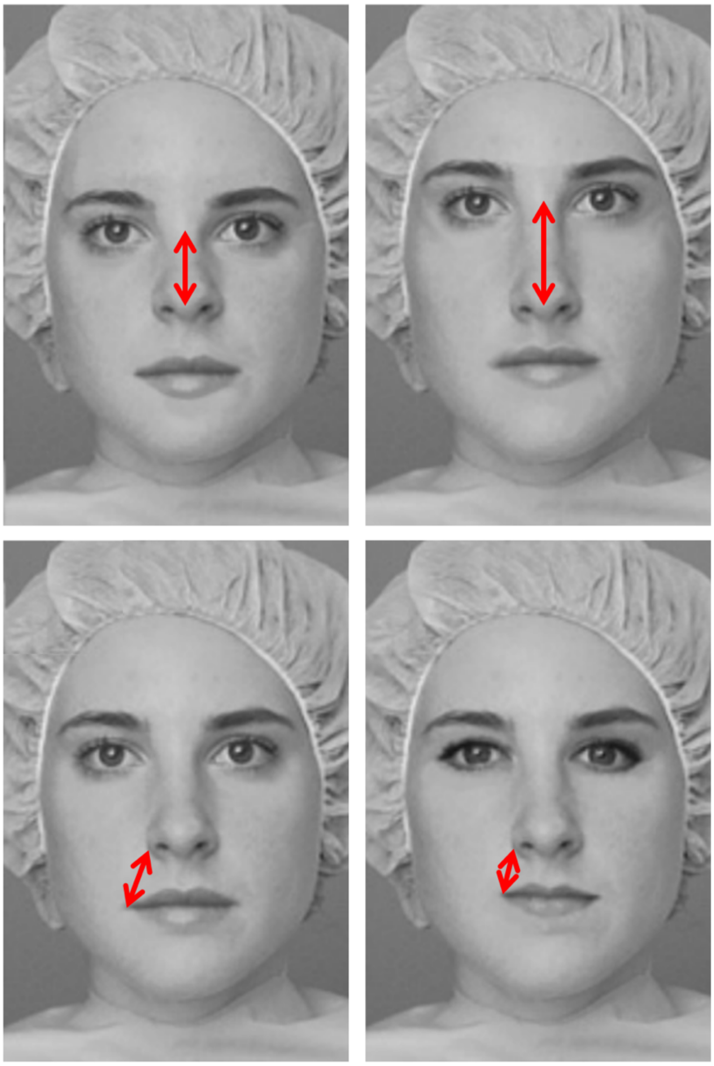


**Figure 1.10.** Stimuli from Mondloch et al. (2002). The top row (a) represents stimuli with configural alterations (eyes and eyebrows moved further down the face in the left image, further up the face on the right image); however, this also produces featural alterations, the nose is bigger/longer in the right image than the left. The bottom row (b) shows featural changes (different eyes and mouths), however, this also alters the configural information, for example the metric distance between the corner of the nose and mouth differs between the left and right image.

Although the configural hypothesis is the leading theory of the cause of the FIE, it has not gone unchallenged. Burton et al. (2015) laid out multiple arguments against the configural account. One issue they raise is that configural theories are underspecified; the configural account revolves around the perception of second order configural cues, the metric distances between key features, however, it does not operationalise this. For example, it does not state how many measurements are made, where to measure from, e.g. the edge, corner or centre of a feature, which distances are the most important for recognition etc. Perhaps the biggest problem Burton et al. (2015) raise with the configural account is that you *cannot* recognise faces based on second order configural cues. The essence of the configural hypothesis is that configural processing and recognition are inhibited by inversion, therefore we must use configural cues to recognise upright faces. However, as shown in Figure 1.11, there is large within-person variation in metric distances of features and therefore this information *cannot be used to individuate a face*. The impaired configural processing and recognition experienced when perceiving an inverted face is therefore a correlation, not causation. A similar point raised in the paper is that changing second order configural cues does *not* inhibit recognition, a claim that should be true if upright faces were recognised by their precise second order configuration. For example, Hole et al. (2002) stretched familiar faces, doubling the length of the face and majorly distorting the second order configuration, yet recognition was not inhibited. Burton et al. (2015) also point out that upright recognition can be harmed by changes which are not caused by second order configuration changes. Familiar face recognition is very robust, yet image negation harms recognition (Kemp et al., 1990) while leaving the second order configuration intact. These criticisms of the configural processing hypothesis do not refute the fact that configural processing is inhibited by inversion.

However, they do suggest that configural processing is not what is driving face recognition with upright faces, and therefore might not be what is inhibiting recognition with inverted faces.

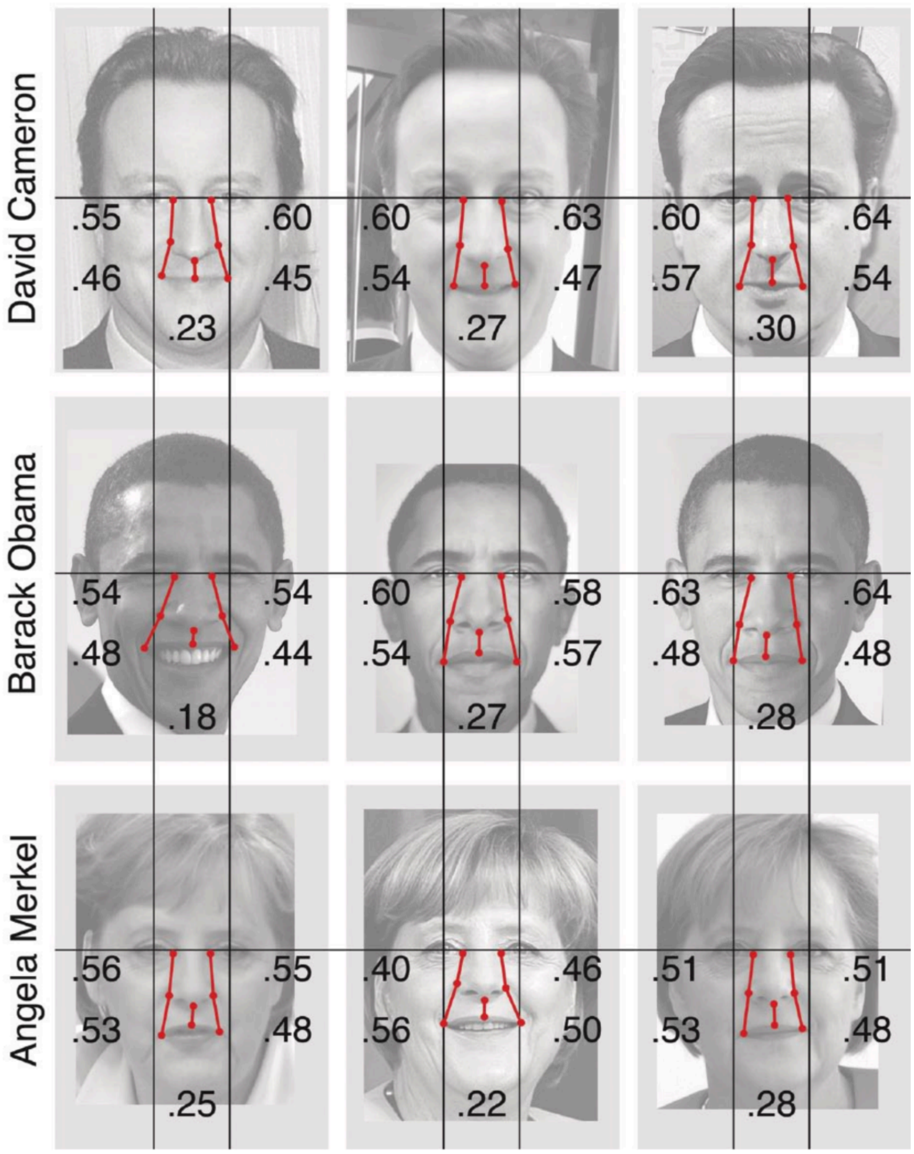



**Figure 1.11.** Figure taken from Burton et al. (2015) demonstrating the within-person variation in measurements between feature distances. The red lines show which distances were measured, and the associated numbers are proportions of the interocular distance.

An alternative account of the FIE is the expertise hypothesis. This theory is similar to the configural account in that it states the FIE is caused by the inhibition of configural processing when perceiving inverted faces; however, it states that this is *not* face specific. Configural processing can be recruited by any stimulus type as long as members of the category share the same first order configuration and therefore are discriminated by their second order configuration, and the viewer has expertise at individuating members of the stimulus type, e.g. bird watchers might be expected to show comparable inversion effects with birds. Diamond and Carey (1986) tested this with dog experts (judges and breeders) and novices (undergrade students). Participants were shown upright and inverted faces and dogs, and at test had to indicate which image in a pair they had seen before. They found that both dog experts and novices showed large inversion effects for faces, but only dog experts showed large inversion effects for dogs. This supported their idea that the FIE is caused by impaired configural processing, and that the effect is not specific to faces but to categories of expertise.

Mixed evidence for the expertise hypothesis comes from the artificial stimuli “Greebles”. Greebles are novel objects with protruding parts which all share a first order configuration and so must be distinguished by their second order configuration. Participants are typically trained to become Greeble experts and can recognise individual Greebles, as one can recognise individual faces (Gauthier & Tarr, 1997). They found that lab-trained Greeble experts were better at recognising upright Greeble parts when shown in the original Greeble (whole) as compared to in isolation (part), similar to the part-whole effect found for faces. However, unlike faces, there was no cost of inversion for Greeble experts or novices in either the studied configuration, transformed configuration or isolated parts condition, casting doubt on the expertise hypothesis explanation of the FIE. Furthermore, the validity of Greebles as good control stimuli for faces has also been questioned. For example, faces vary far more widely than Greebles, and we all have a lifetime of experience with faces, but Greeble experts receive just ~10-15 hours of training. Furthermore, faces move for speech and expressions, however, Greebles have no non-rigid deformations. Moreover, in Ashworth et al. (2008) the faces are cropped to remove most of the external features, however, they state that Greebles cannot be cropped similarly as this would involve cropping out the distinguishing features. However, the external features of unfamiliar faces are known to be distinguishing (Bonner et al., 2003) yet the crop is applied to faces.

As I have noted throughout this face inversion effect section, the face stimuli used in experiments are rarely representative of how we see faces. Face stimuli crop most of the person out (the body), and often crop out the external features and background. Apart from moving away from using line drawings, the ecological validity of stimuli has not progressed over the years, as demonstrated in Figure 1.12. The use of such stimuli has not been challenged in the FIE literature and has only been indirectly challenged in the face learning literature (studies examining within-person variation use full colour images and do not crop the face severely; Jenkins et al., 2011). As the face learning literature had to move away from tightly controlled stimuli to discover the true effects, so the face inversion literature might benefit from using less tightly controlled, more naturalistic images.

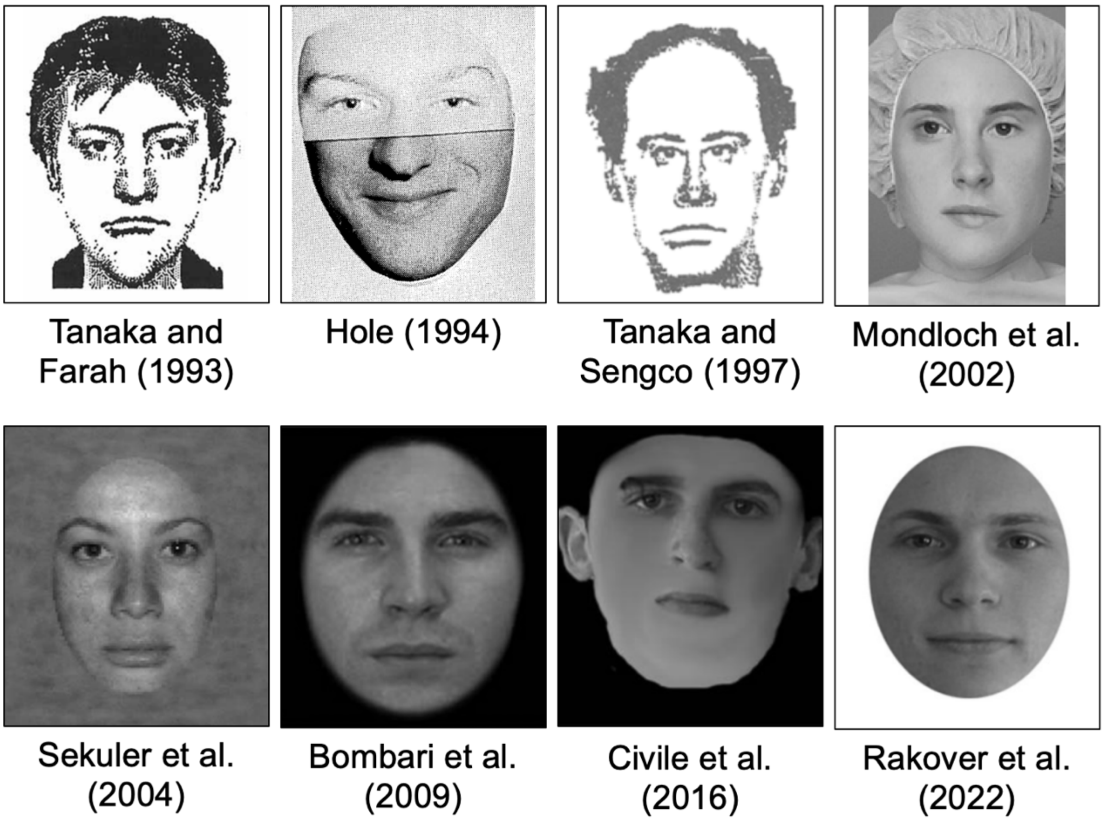



**Figure 1.12.** Example stimuli taken from 8 different journal articles spanning 29 years from 1993 to 2022. The stimuli are not in colour, remove the background, crop the body out, and most remove/occlude the external facial features. Note that the ecological validity of the stimuli has remained fairly stable across time.

Another key issue with all theories of the face inversion effect and most experiments studying the effect, is that unfamiliar and familiar faces are treated as being the same. As mentioned in an earlier section, Burton (2013) outlines how lumping unfamiliar and familiar faces into 1 category of “faces” has been a mistake of the literature which has led to slow progress in the field. Many studies begin by referencing how robust face recognition is, clearly referring to the ability to recognise a familiar face. However, many studies are conducted using unfamiliar faces (Rakover et al., 2022). Papers also often refer to “distinguishing” between faces and “recognising” faces and use these terms interchangeably. However, as mentioned above, the processing of familiar and unfamiliar faces is different. As shown in card sorting tasks (Jenkins et al., 2011) we are excellent at telling familiar identities together (i.e. *recognising* faces), but poor at telling unfamiliar identities together, however, we are excellent at telling unfamiliar identities apart (i.e. *distinguishing* between faces). Treating familiar and unfamiliar faces as one type of stimulus class in experiments and theory may have slowed down progress in our understanding of how faces are recognised (or rather how familiar faces are recognised, and unfamiliar faces are perceived). It might then be fruitful to reexamine the face inversion effect with familiar and unfamiliar faces separately.

In recent years, the face inversion effect itself has been questioned. For example, Gerlach et al. (2023) found comparable inversion effects for objects and faces. They tested the idea that the FIE is found because tasks generally require within-category discrimination of faces and objects, however, in day to day life we only do this with faces, not objects. They compared a within-category discrimination face task with an object decision task. Participants judged whether upright or inverted stimuli were real objects or not in an easy condition (non-objects) and a hard condition (chimeric non-objects made up of real object parts). Participants also viewed upright faces then completed an old/new face memory task with upright and inverted face stimuli. For the difficult object task and the face old/new task, they found no effect of inversion (accuracy reductions <1%), however, they found a small but significant inversion effect for the easy object task. All conditions suffered similar inversion effects in the reaction times. In order to control for task differences, in Experiment 3, participants completed old/new memory tasks with objects and faces. They found no inversion effects in the accuracy data for objects or faces. Furthermore, they found comparable inversion effects in the RT data for faces and objects. Across 3 experiments, they found no differences in the effect of inversion between faces and objects. This paper argues we are experts at differentiating faces when upright and therefore the face inversion effect is actually a face upright effect.

Further lack of evidence for the face inversion effect comes from Rezlescu et al. (2016). Participants completed both decision and within-category discrimination tasks with faces and cars. For the decision tasks, participants saw 1 intact and 2 scrambled Mooney faces and had to indicate which one was a face. An equivalent task was completed with cars. For the within-category discrimination tasks, participants had to match a target image to one of three images at a different viewpoint. For the decision task, they found equally large inversion effects for faces (27.1%) and cars (27.8%). Similarly, in the within-category discrimination tasks, they again found large and equal inversion effects for faces (25.1%) and cars (27.5%). These results suggest that faces are not special in the sense that they do not suffer disproportionately from inversion. They also found evidence that the large inversion effects found are *not* linked with expertise. When they removed the top performing participants from the car individuation task, the results did not change, suggesting that inversion effects do not vary with expertise. Similar results were found again by Gerlach and Mogensen (2023). They found that when performance on inverted trials was regressed out of performance on upright trials, a more reliable measure than calculating difference scores, they found equal-sized correlations between performance on the Cambridge Face Perception Test (CFPT) and Cambridge Face Memory Test (CFMT), and between performance on the CFMT and the Cambridge *Car* Memory Test (CCMT). They did find a larger decrease in accuracy for inverted faces compared to inverted cars. The authors explain that this is troubling for an account of face perception which states *only* faces are processed holistically. This instead suggests that faces and objects may be processed via a shared mechanism, perhaps holistic processing, however, it might be more important to face processing.

To summarise the literature reviewed, there are different brain regions which are face-sensitive, and they form networks for the efficient processing of faces. However, whether faces are processed in a unique way or not and therefore whether they are a special class of stimuli is still debated. Familiar faces may be regarded as special, as it has been demonstrated familiar face recognition is more accurate and quicker than for unfamiliar faces, due to their underlying stable face representations. There is a large amount of variation not just between identities but particularly within identities, and it is this ability to tolerate within-person variation and tell a face together which defines familiar face recognition. For these stable face representations to be formed, unfamiliar faces must be learnt and compiled together into 1 representation. The key factors of face learning seem to be incorporating within-person variation into learning and top-down information to aid in linking these variable instances of faces together. Due to our exceptional performance with upright faces and the large decreases in accuracy reported for inverted faces, multiple theories have been suggested to account for the face inversion effect. Mixed evidence exists for the holistic, configural, and expertise hypotheses. However, a potential issue with much of the face inversion literature lies in the artificial stimuli used – they are often unfamiliar, with the body, background, and external facial features cropped out.

Based on the reviewed literature, my thesis has 2 aims: to explore the role of top-down binding cues in face learning, and to reexamine the face inversion effect using naturalistic stimuli and investigate if face familiarity has a modulating effect.

Firstly in chapter 2, I investigated face learning. As discussed above, Ritchie and Burton (2017) showed that learning to recognise initially unfamiliar identities from more variable photographs leads to more accurate subsequent recognition. However, it is known that we are poor at tolerating within-person variation of unfamiliar faces (Jenkins et al., 2011). Therefore, across 4 experiments in chapter 2, I sought to investigate whether top-down binding cues might play a role in face learning by potentially cohering together highly variable images of initially unfamiliar faces into a stable face representation which would then aid subsequent recognition. As noted throughout the present chapter, ecological validity of stimuli is important in experiment design. Therefore, in chapter 2, while the images were cropped to include primarily the face as the experiment focused on within-person variation of faces, the upper body, external features and backgrounds were deliberately not cropped out, to ensure that the stimuli were as realistic as possible while removing the majority of the body.

In chapter 3, I investigated the face inversion effect. As discussed above, Yin (1969) showed that faces suffer larger inversion effects than other non-face stimuli. However, the stimuli used to study the FIE are not ecologically valid and are often heavily cropped, removing the body, neck, background and sometimes external facial features. Across 6 experiments, I investigated whether image cropping plays a role in the FIE by using cropped stimuli typical of previous experiments (body and background removed), or more naturalistic ecologically valid images typical of our day-to-day viewing experiences (body and background included). Also, as mixed evidence has been found regarding the role of familiarity on the FIE, I also tested both familiar and unfamiliar faces to see if face familiarity played a modulating role.

# Chapter 2 – Face Learning

## Introduction

Throughout our lives, we learn to recognise new faces, for example when we meet new people at work or at a party. However, the means by which we learn to recognise a face is still poorly understood. This is surprising as the large differences between familiar and unfamiliar face perception are well documented (Johnston & Edmonds, 2009; Young & Burton, 2017). Faces have been shown to vary greatly even within an individual (Jenkins et al., 2011), and familiar face recognition can withstand such variation, however, unfamiliar face perception cannot, and in fact is more closely image-bound. For example, we are able to recognise familiar faces across a range of viewing conditions, and even when images are distorted, stretched or blurred (Burton et al., 1999; Hole et al., 2002; Lander et al., 2001). This robust ability to recognise a familiar face across changes is different from our recognition of recently learnt unfamiliar faces. In an old-new recognition task, Bruce (1982) found that unfamiliar faces were recognised less accurately compared to familiar identities. Furthermore, unfamiliar face recognition is closely bound to the specific images used in an experiment, and recognition accuracy decreases as the difference between study and test images increases (Longmore et al., 2008). The experiments in this chapter aim to address the question of how faces are learnt, by exploring how top-down information may be used in the face learning process.

One of the challenges with face learning is that individual faces can vary greatly in their appearance (Jenkins et al., 2011). For example, there are short term changes to the face caused by rigid movements such as head nodding/shaking and viewpoint, and non-rigid movements such as speech and emotional expressions (Hunnisett &

Favelle, 2021; Lander et al., 1999; O'Toole et al., 2002). Faces also vary over longer time periods, for example different hairstyles, make-up, health, aging, environmental conditions etc. (Hassaballah & Aly, 2015; Jacob et al., 2010). This within-person variation is, to a degree, idiosyncratic – that is, some of the dimensions an individual face varies on are different to the dimensions of another face (Burton et al., 2016). This therefore means that learning to recognise one face does not generalise and aid recognition of a different identity. One must learn how an individual varies in order to become familiar with them and recognise them.

For a given unfamiliar face, one has not observed the range of how the face can vary. Therefore, in matching tasks, the viewer is not relying on the person representation containing the observed within-person variation, but instead might use less sophisticated image matching strategies (Hancock et al., 2000). This is in comparison to matching images of familiar identities, where both images of the individual are easily recognised and therefore accurately matched. These differences between familiar and unfamiliar faces explain why unfamiliar face matching is surprisingly poor and familiar face matching is highly accurate (Bruce et al., 1999, 2001). Matching tasks are often used in face perception research as they are sensitive to the range of familiarity, and they rely on perception not memory (Ambrus et al., 2021; Clutterbuck & Johnston, 2004).

In this chapter, I investigate how it is possible to learn to recognise a face. One key finding is that observing within-person variation is known to aid face learning. For example, in a study by Ritchie and Burton (2017) participants learnt to recognise previously unfamiliar faces by viewing images of the faces blocked by identity with their name presented above. The to-be-learnt identities were either depicted in highly variable photographs, e.g. photos taken on different days with different lighting, viewpoints, make-up etc., and low variability photographs, e.g. stills from an interview. Participants then completed a matching task with novel images of the identities to test learning. They found that faces learnt from highly variable photographs were matched more accurately compared to identities learnt from low variability photographs. Furthermore, when matching images of a target identity amongst a large array of distractors, matching accuracy increases as the number of images of the target increases (Bindemann & Sandford, 2011; Dowsett et al., 2016). Moreover, within-person variation has been shown to benefit face learning independently of viewing time. Murphy et al. (2015) showed participants either the same 6 photographs or new photographs of an individual on each trial during a learning phase, with exposure time held constant across the two conditions. Participants were more accurate at matching identities learnt from a range of images compared to identities learnt from repeated images. Together, these studies demonstrate that observing within-person variation aids face learning and subsequent recognition.

However, as noted above, unfamiliar face recognition is less able to tolerate within-person variation. For example, Jenkins et al. (2011) gave participants printed photographs of 2 similar-looking celebrities (20 variable images of each face) to sort into identity piles, 1 pile of photos per identity. Participants unfamiliar with the celebrities split the images into a mode of 9 piles, perceiving there to be 9 different identities, not 2. Participants were unable to tolerate the within-person variation and tell the images together. However, when thinking about face learning, a problem arises: how is the viewer, who is initially unfamiliar with the face and unable to tolerate within-person variation, able to cohere multiple variable images together? How does the cognitive system know which images to cohere together, and how does it incorporate variable instances of the face into a face representation?

In day-to-day life when recognising people, we do not rely on visual information from the face alone, but rather we recognise people from their face, voice, body, location etc. (Kamachi et al., 2003; Yovel & O'Toole, 2016). It may be these multiple sources of top-down context information which help to cohere within-person variation into a representation. When speaking to an unfamiliar person for the first time, one observes within-person variation such as different viewpoints and expressions, yet it is clear that one is speaking to just one individual and this variable visual face information should be cohered together into a representation. Furthermore, when meeting people across encounters, top-down context information again may help to cohere together visual information into a face representation. For example, if you have a new work colleague and you enter your place of work on their second day, you know to expect that the person sat in their desk is your new colleague. Or for example when watching the news, the name of the person about to appear on screen may be announced and it is presented on the screen. These cues in day to day life tell us to bind together the variable instances seen into one representation.

Behavioural studies suggest that top-down information aids image coherence. In a card sorting task, Andrews et al. (2015) presented participants with a deck of printed photographs of faces, 20 images each of 2 similar looking identities. Half of the participants were told that there were 2 identities in the photographs (two-sort condition), and half were not given any top-down information (free-sort condition). Participants were instructed to sort the photographs into identity piles, 1 pile per identity. They found that the free-sort group made a mean of 6.8 piles, in comparison to the two-sort group who were instructed to make 2 piles. Both groups made minimal misidentification errors, rarely putting multiple identities into 1 pile. This replicated the original findings of Jenkins et al. (2011) that unfamiliar viewers cannot tolerate within-person variation and therefore make many piles, and extended the findings, that providing minimal top-down information gave viewers the ability to cohere together the correct variable images.

Further evidence comes from Schwartz and Yovel (2016), where participants were familiarised with faces, then completed an old-new recognition task. In the learning phase, participants either saw identities represented by 1 image repeated 10 times with a name (name condition), 10 images which varied in rotation and lighting (multiview condition), or 10 images which varied in rotation and lighting with a name (multiview + name condition). In the follow-up matching task, they found that participants were significantly more accurate when the 10 variable images were paired with a name compared to without. The authors suggested that the "name labels may link the different images to a common view-invariant representation" (Schwartz & Yovel, 2016, p. 1500). However, the amount of variation in the multiview conditions, changes in viewpoint and lighting, was much less than the levels of within-person variation experienced in day to day life across encounters. Further research with more variable images is needed to investigate if name labels can cohere highly variable images together.

Computational models also lend support to the idea that top-down information supports image coherence and face learning. For example, Kramer et al. (2018) developed computational models using PCA and LDA. They computed statistical descriptions of over 4000 images of 335 IDs by performing PCA, and they created a range of familiarity by including 1 to 159 images per person. They trained various models by performing PCA + LDA on the images, which produced a space which could discriminate between the 335 identities, and found that the model was able to replicate some key face recognition findings, for example the higher the familiarity (e.g. the more training images of an ID), the better the recognition accuracy of the model. To test the effect of top-down information (LDA) on the model, they randomly selected 80 images of Ryan Reynolds to be training images and 20 to be test images. At the PCA stage, the 80 training images were included, and at the LDA stage the 80 training images were either left in (trained identity condition) or left out (untrained identity condition). The 20 test images were then projected into the space and the Euclidian distance was calculated between each combination of pairs of images of Ryan Reynolds. They found that the mean Euclidian distance was closer in the trained identity condition compared to the untrained condition, demonstrating that top-down information reshaped the underlying space to cluster images of the same identity together. The authors concluded that, "Understanding how faces become familiar appears to rely on both bottom-up statistical image descriptions (modelled here with PCA), and top-down processes that cohere superficially different images of the same person (modelled here with LDA)" (Kramer et al., 2018, p. 46).

Across four experiments in this chapter, I investigated whether top-down information is necessary for face learning. In the first experiment, an attempt was made to replicate the experiment and results of Ritchie and Burton (2017). Participants completed a face learning task, whereby they viewed high or low variability images of identities with top-down information (blocked by identity with the identity's name presented above). As in Ritchie and Burton (2017), this was followed by a matching task to test face learning. It was hypothesized there would be a learning benefit to observing high variability images compared to low variability images. The aim of Experiment 2 was to test whether removing top-down information from the learning phase (by removing the names and unblocking the design) using the same images as in Experiment 1 would remove the advantage of observing highly variable images.

Experiments 3 and 4 shared the same aim as Experiments 1 and 2 but offered improved experimental design.

# Experiment 1: Face learning with top-down information

## Introduction

In this first experiment, the aim was to replicate the findings of Ritchie and Burton (2017), that identities learnt from high variation images will be better recognised than identities learnt from low variation images. In order to achieve this, top-down information was included in the learning phase to aid face learning, by serving as a cue to bind together the variable instances seen of each face. It was hypothesized that when identities are learnt with top-down information, high variation identities will be better recognised than low variation identities.

## Methods

### *Participants*

As this experiment (and all experiments in this thesis) had multiple within-subjects factors, a power analysis could not be carried out using G*Power to determine the number of participants required to find the effect. This experiment aimed to replicated Experiment 2 of Ritchie and Burton (2017) which used 30 participants, therefore a larger sample of 60 participants was collected to ensure sufficient power.

54 right-handed participants (25 females, 29 males, mean age = 25.97, SD = 5.38, 1 participant gave an inaccurate DOB and therefore their data is not included in the mean age or SD) took part in this experiment. 6 additional participants were excluded from the analysis due to data recording issues. Participants were recruited

from Prolific and were paid £4 to compensate their time. Informed consent was gained, and the experiment was approved by the University of York Psychology Ethics Board.

***Materials***

44 full-colour photographs of 10 foreign celebrities (or ‘targets’, 5 female) were collected from the internet for use in this experiment (440 in total). Foreign celebrities were used as the same type of stimuli were used in Ritchie and Burton (2017), identities were unknown to participants, and multiple images of each identity were easily accessible to use in both the low and high variation conditions. 34 of the images of each celebrity were high variability images which were in full colour and were 3/4 to full face. Apart from these constraints, they varied greatly in terms of personal variation, e.g. different facial expressions, make-up, hairstyles and age. They also varied greatly in non-personal variation, e.g. different camera angles, lighting, distance from the camera (see Figure 2.1.1). As in Ritchie and Burton (2017), 10 of these images were used as high variation images for the learning task, the remaining 24 high variation images were used in the matching task, 16 images to form the match trials, and 8 images were paired with foils to form the mismatch trials. The low variability images (10 images per identity) were stills taken from videos. These images varied somewhat in personal variation, e.g. different facial expressions and viewpoints, and varied minimally in non-personal variation, e.g. same background and lighting (see Figure 2.1.1, and see Appendix A for further examples).


**Figure 2.1.1.** Example stimuli. The top row depicts high variability images of a target, and the bottom row shows low variability images of the same target.

Each target was paired with a foil identity who resembled them for use in the matching task (Andrews et al., 2015; Murphy et al., 2015). Images of foil identities were found by searching for foreign celebrities who matched verbal descriptions of the target identities (e.g. female, short brown hair, mid 30s) and was checked by the author LRS that they bore a strong resemblance to the target identity. For example, Figure 2.1.3 shows a target (right) paired with a highly similar looking foil (left). 8 high variation images were collected of each foil for use in the matching task, as in Ritchie and Burton (2017). The images of the learned identities in the matching task were high variation novel images not seen during the learning task, in order to truly test face recognition – the ability to generalise and recognise a novel image – as opposed to using a low variation image would closely resemble the images seen during learning, or as opposed to using the same image, and testing image recognition – the different and easier task of recognising the same image, rather than recognising the identity. Furthermore, images of 40 novel identities and 40 associated foils were collected for the matching task to provide a baseline of face matching ability. 3 images of each novel identity and 1 image of each foil were collected.

The stimuli were cropped around the head at a ratio of 380:570 pixels and included the remaining background. As noted in chapter 1, ecological validity of stimuli is important in experiment design. The images were cropped to include the whole head and upper body and the background was not removed. This was done in order to use a similar crop as in Ritchie and Burton (2017), and it also maintained some ecological validity (the images were realistic looking as the upper body / neck / external features were not removed). The experiment was built using PsychoPy3 (https://pavlovia.org/; Bridges et al., 2020; Peirce et al., 2019) and run online at Pavlovia. Photographs were displayed in 0.4 x 0.5 'height' units, in order that the images would scale appropriately to each participants' screen as this experiment was run online.

### ***Design and Procedure***

#### *Learning Phase*

In the learning phase, participants saw 10 photographs of each identity sequentially. The same top-down binding cues as used in Ritchie and Burton (2017) were used here: the photographs were blocked by identity and the name of the identity was presented throughout to ensure that participants understood all of the images being presented to them in a block were of the same identity. Participants saw 5 of the identities represented by high variation images and 5 represented by low variation images; this was counterbalanced across participants and produced 100 trials. At the start of a trial, a fixation cross appeared in the centre of the screen for 0.5s with the identity's name presented at the top of the screen on a grey background. The face was then displayed for 5s, and the name remained at the top of the screen, as shown in Figure 2.1.2. The order of photographs within the blocks, and the order of blocks was randomised for every participant. To ensure participants attended to the stimuli, they were instructed to indicate via a button press (1 or 0) whether the eyes were looking at the camera or not. This attention task was used as it is orthogonal to the experimental task, it would provide a measure of attention which is unrelated to the experimental data, a requirement of collecting data on Prolific. Furthermore, when learning faces, the eyes are often attended to frequently and for longer durations and may therefore aid the face learning process (Sekiguchi, 2011). Participants were on average 94.0% accurate. The image remained on screen for the full 5s even after the response was recorded to ensure the exposure was identical for all participants. Therefore, the learning phase lasted 9 mins 10s.

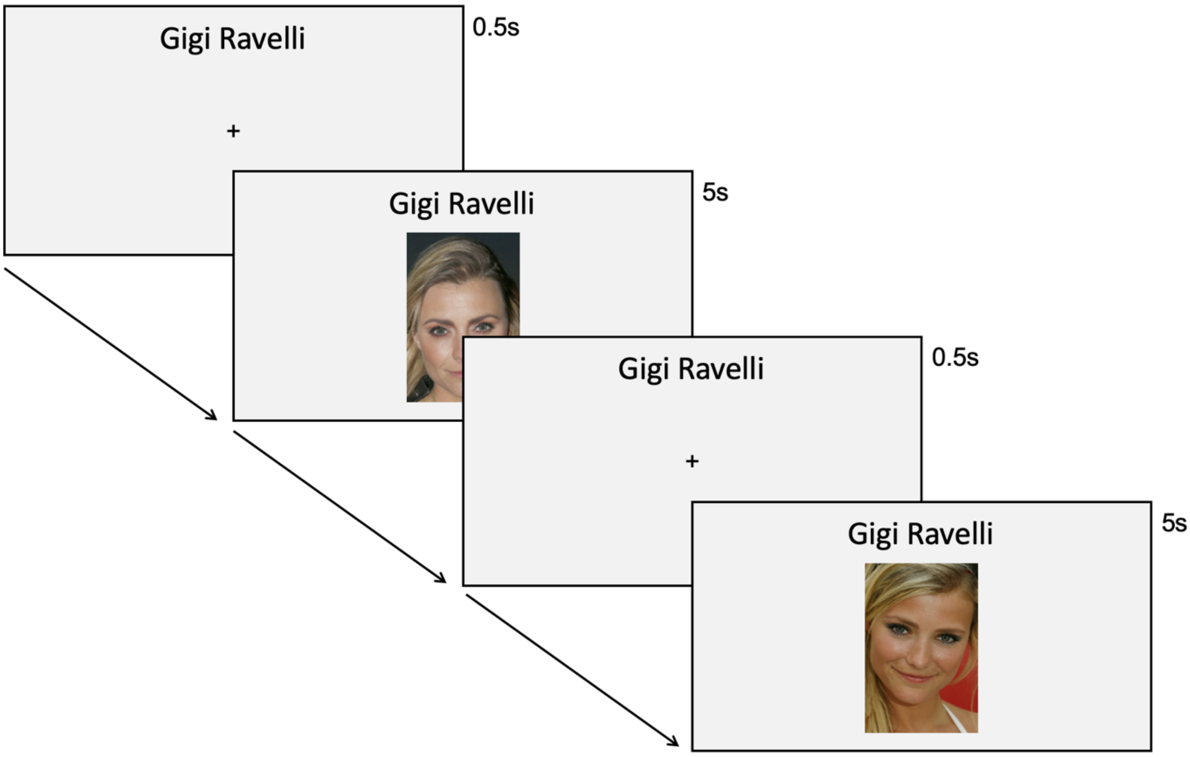



**Figure 2.1.2.** Example trials from the learning phase of Experiment 1. Images of each identity were shown in identity blocks with their name presented at the top of the screen throughout.

*Matching Task*

To test learning, participants completed a matching task with novel images of the identities shown in the learning task and novel identities. In the matching task, on each trial participants saw 2 images side by side on the screen and indicated via a button press (1 or 0) whether they depicted the same person or different people; the response keys were counterbalanced across participants. Trials began with a fixation cross presented in the centre of the screen for 0.5s followed by the pair of faces, as shown in Figure 2.1.3. The images remained on screen until participants made a response. Half of the trials were match trials and half were mismatch. There were 8 match and 8 mismatch trials for each learned identity (80 trials for low variation identities and 80 trials for high variation identities), and 1 match and 1 mismatch trial for each novel identity (80 trials in total). These were included to provide a baseline of face matching accuracy. The order of trials was randomised for each participant. Participants were asked to respond as quickly and accurately as possible. 20 attention trials were added to ensure participants were attending to the task. Attention trials consisted of identical image pairs (10 male, 10 female) where participants would be expected to display ceiling effects. Participants who scored below 80% were removed from subsequent analyses; participants were on average 98.4% accurate.

This experiment had 2 independent variables, one with 2 levels and one with 3 levels, producing 6 conditions. The familiarity level was either novel, low variation identity, or high variation identity. The trial type was either match or mismatch. Matching accuracy and RTs were recorded. On average, the experiment lasted approximately 33 minutes.

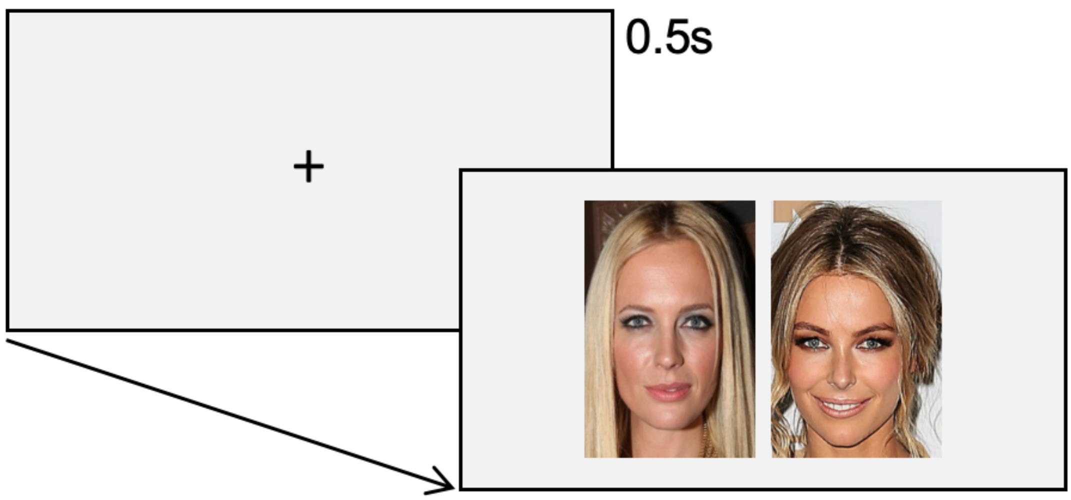



**Figure 2.1.3.** An example trial from the matching task.

**Results**

Match and mismatch trials were analysed separately as in previous studies (White et al., 2014; Ritchie & Burton, 2017). It is suggested that different cognitive processes underly match and mismatch trials, and therefore different patterns of results may be found by analysing match and mismatch trials separately (Megreya & Burton, 2007). As faces vary idiosyncratically, one must learn how an individual face varies in order to be able to later recognise them. As discussed in Ritchie and Burton (2017), participants learn to recognise the identities during the learning phase without it leading to a liberal accepting of similar looking identities. Therefore, in the matching task, on match trials, participants may recognise each photo of the learned identity, participants tell them together, and produce high accuracy and fast reaction times. However, on a mismatch trial, as one of the faces is unfamiliar, a different process may take place compared to in the match trials, for example image matching strategies to tell the identities apart.

*Match trials: Accuracy*

Match trials of faces learnt from high variability photographs produced the highest accuracy, followed by low variation IDs, followed by novel IDs as shown in Figure 2.1.4. A one way repeated measures ANOVA (factor: familiarity, levels: novel, low variation, high variation) was performed on the match accuracy data. The data violated the assumption of sphericity ($p$ = .003) therefore a Greenhouse-Geisser correction was used. A significant main effect of familiarity was found ($F(1.67, 88.29) = 24.34$, $p < .001$, partial $\eta^2 = 0.32$). Bonferroni corrected pairwise comparisons revealed that identities learnt from low variation photos ($p < .001$), or high variation photos ($p < .001$) were matched significantly more accurately than novel identities.

There was no significant difference in matching accuracy for low or high variation ($p$ = .282).

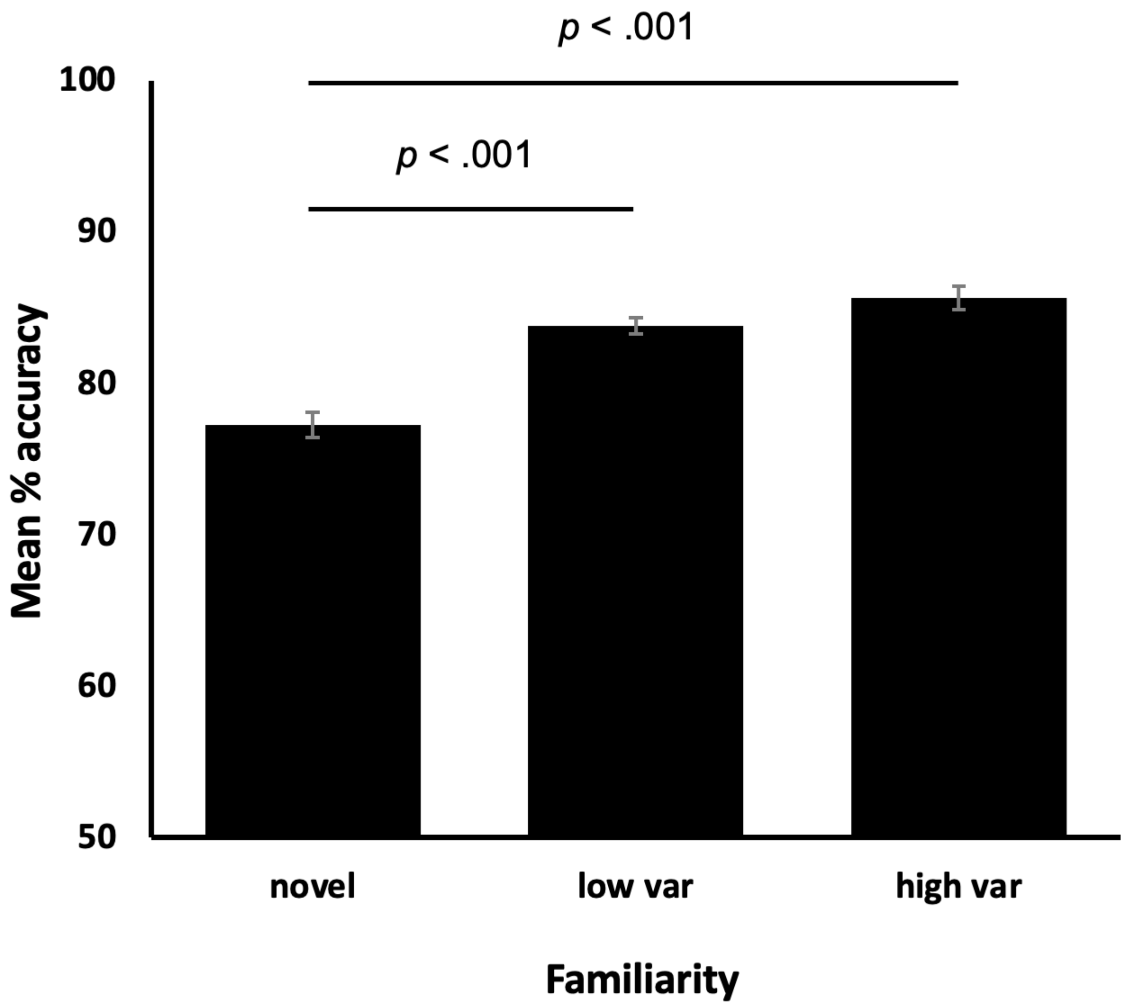



**Figure 2.1.4.** Mean accuracy for match trials for novel identities, identities learnt from low variation images and identities learnt from high variation images with top-down information. Here and throughout, error bars show within-subject standard error of the mean calculated by the method presented in Cousineau (2005).

*Match trials: RTs*

Faces learnt from high variability photographs were matched the fastest, followed by low variation IDs, followed by novel IDs, as shown in Figure 2.1.5. A one way repeated measures ANOVA was performed on the match trials RT data. The data violated the assumption of sphericity ($p$ = .010) therefore a Greenhouse-Geisser correction was used. A significant main effect of familiarity was found ($F(1.72, 91.28) = 27.92$, $p < .001$, partial $\eta^2 = 0.35$). Bonferroni corrected pairwise comparisons revealed that identities learnt from high variation photos ($p < .001$), or low variation photos ($p < .001$) were correctly matched faster compared to novel identity matches. High variation identities trended towards being matched faster than low variation identities, however, this was not significant ($p = .057$).

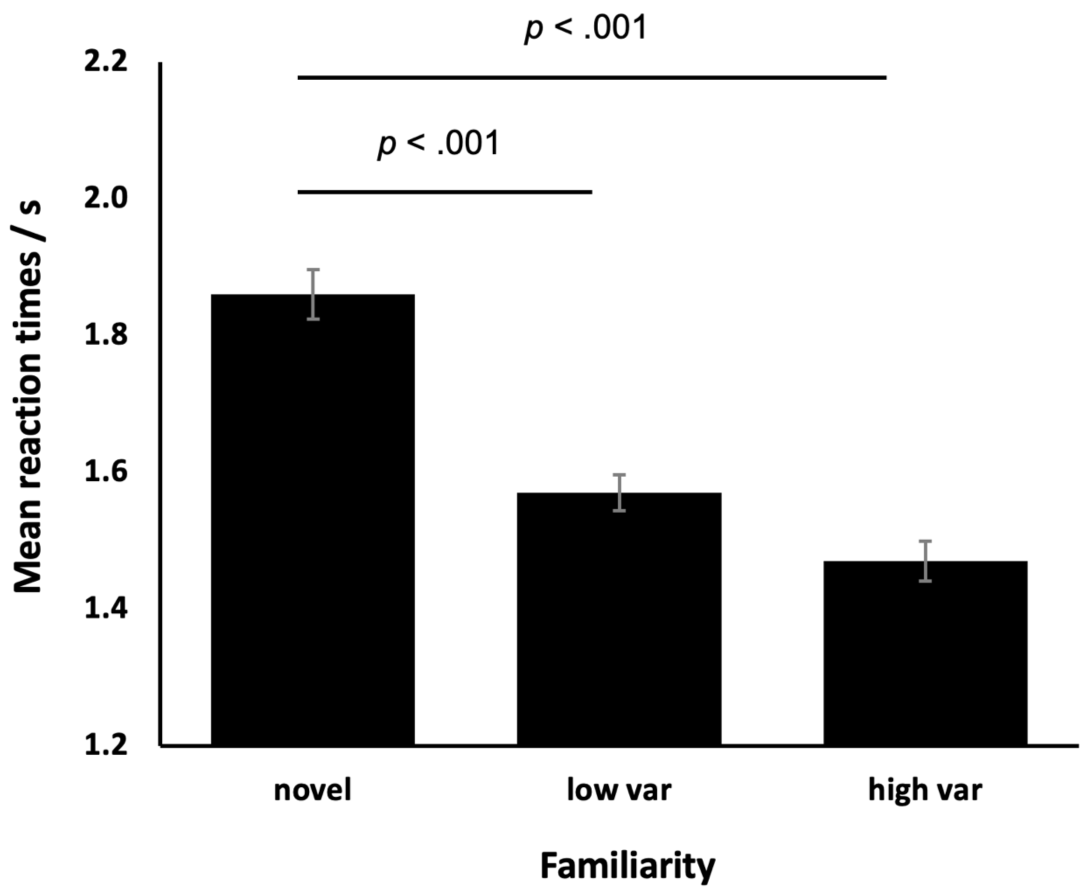


**Figure 2.1.5.** Mean reaction times for match trials for novel identities, identities learnt from low variation images and identities learnt from high variation images with top-down information.

*Mismatch trials: Accuracy*

Mean accuracy for mismatch trials is shown in Figure 2.1.6. A one way repeated measures ANOVA was performed on the mismatch trials accuracy data. The data did not violate the assumption of sphericity ($p$ = .361). No significant main effects of familiarity were found (F(2, 106) = 0.85, $p$ = .430, partial $\eta^2$ = 0.02).

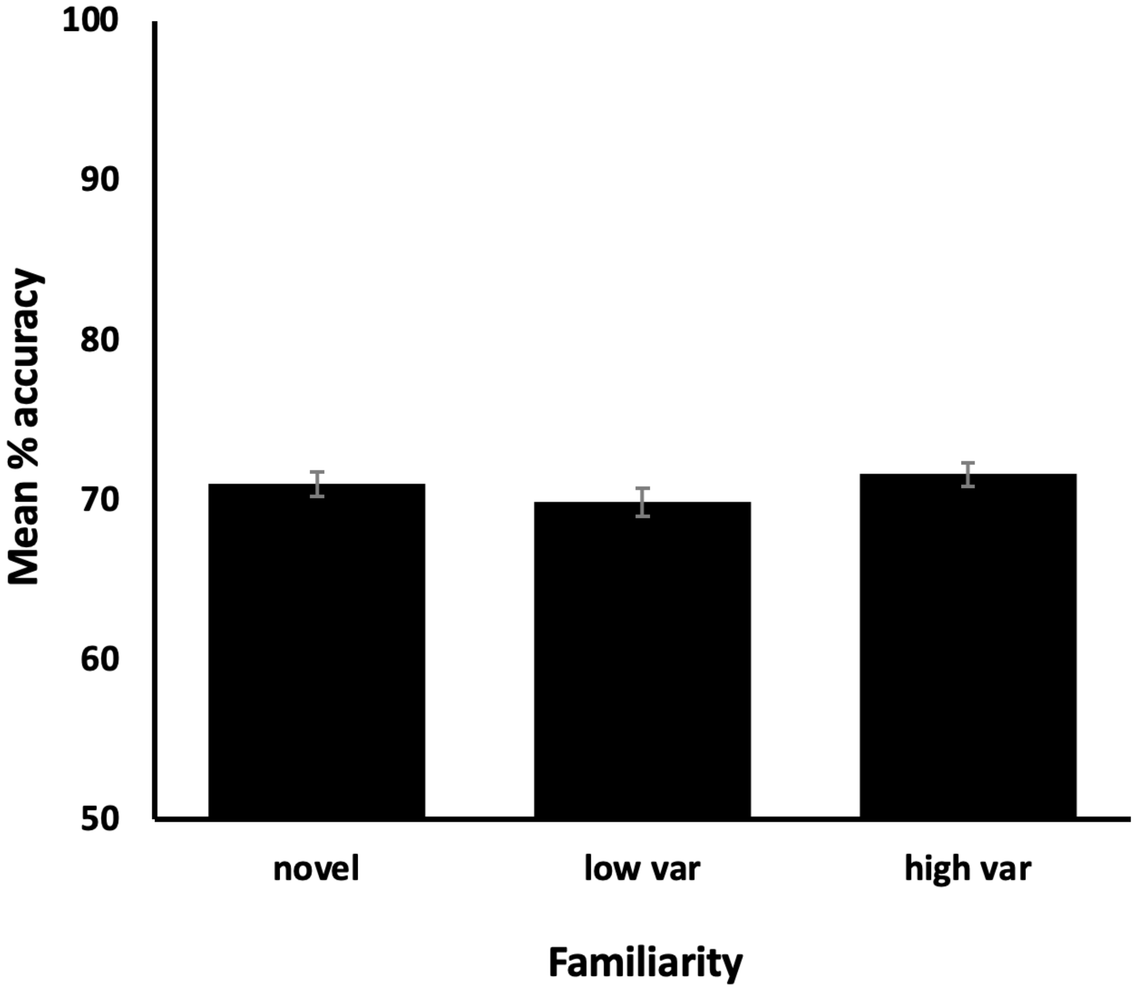


**Figure 2.1.6.** Mean accuracy for mismatch trials for novel identities, identities learnt from low variation images and identities learnt from high variation images with top-down information.

*Mismatch trials: RTs*

Mean RTs for mismatch trials is shown in Figure 2.1.7. A one way repeated measures ANOVA was performed on the mismatch trials RT data. The data violated the assumption of sphericity ($p$ = .044) therefore a Greenhouse-Geiser correction was applied. No significant main effects of familiarity were found (F(1.80, 95.21) = 1.18, $p$ = .309, partial $\eta^2$ = 0.02).

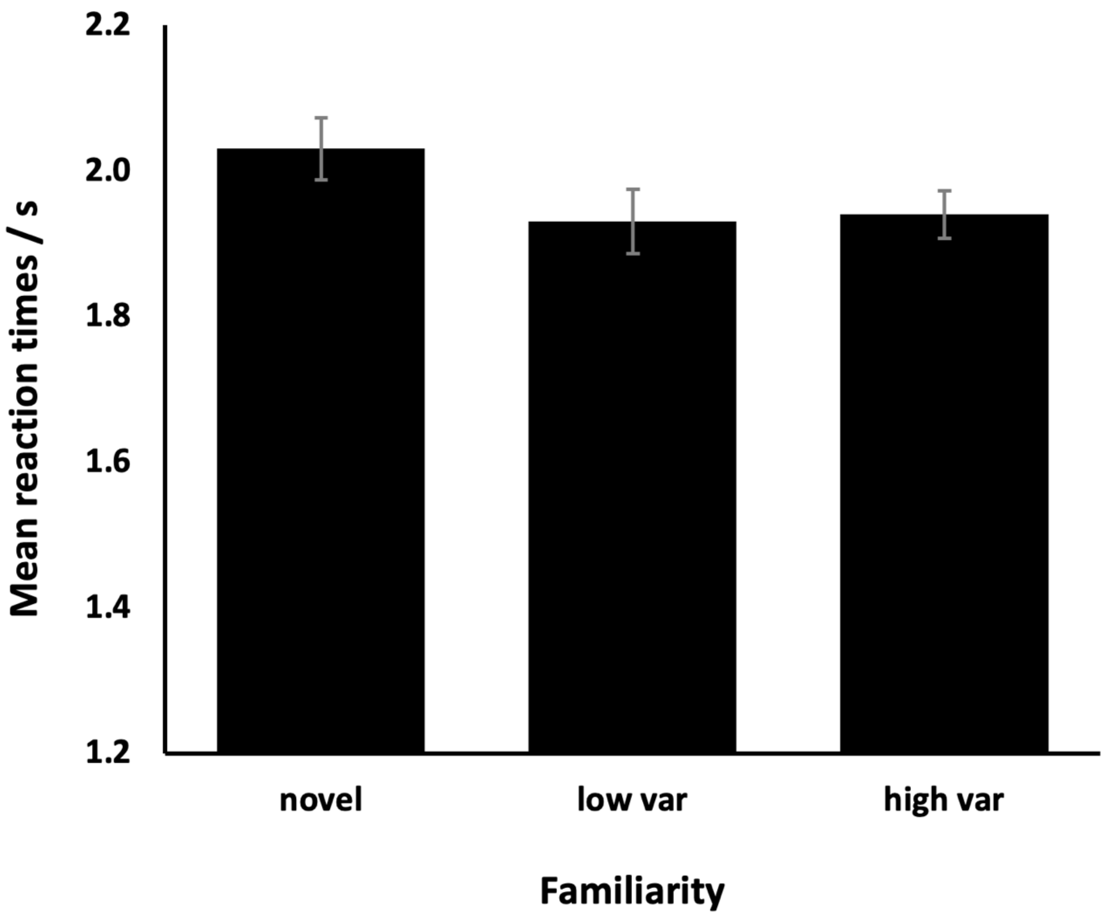


**Figure 2.1.7.** Mean reaction times for mismatch trials for novel identities, identities learnt from low variation images and identities learnt from high variation images with top-down information.

**Discussion**

In this experiment, the pattern of results found here is the same as the pattern found in Ritchie and Burton (2017), however it failed to reach significance. A benefit to learning identities from high variation images in match trials was found compared to identities learnt from low variation trials. Identities learnt from high variation images were both matched more accurately and faster than identities learnt from low variation images. This is likely because participants were able to build robust face representations, cohering together even highly variable images. This robust face representation was able to generalise to recognise the novel images in the matching task, producing the higher and faster performance for high variation identities.

In the following experiment, top-down information was removed from the face learning process. The same images were used as in Experiment 1, however no information was given to participants about which images to cohere together. It was predicted that participants would not be able to cohere the highly variable images together and therefore would produce similar matching accuracy for identities learnt from low or high variation identities.

# Experiment 2: Face learning without top-down information

## Introduction

The following experiment aimed to reduce the effect found in Experiment 1. Although only approaching significance, the previous experiment showed the same pattern as that of Ritchie and Burton (2017), that identities learnt from high variation images were more accurately and more quickly matched than identities learnt from low variation images. Having established the same pattern, the aim of the following experiment was to conduct a similar experiment but to remove the top-down information during learning in order to try and remove the advantage of observing more within-person variation during face learning. It was hypothesized that there would be no differences in the accuracy or reaction times between the identities learnt from low or high variation images.

## Methods

### *Participants*

54 right-handed participants (25 females, 29 males, mean age = 26.48, SD = 6.74) took part in this experiment. 6 additional participants were excluded from the analysis (3 participants experienced data recording issues, 2 participants failed the attention check in the matching task, and 1 participant had missing cases in their RT data). Participants were recruited from Prolific and were paid £4 to compensate their

time. Informed consent was gained, and the experiment was approved by the University of York Psychology Ethics Board.

***Stimuli, Design and Procedure***

The same stimuli were used as in Experiment 1. The design and procedure were the same as in the previous experiment apart from 1 key manipulation – top-down information was removed. The top-down binding cues implemented in Experiment 1 were the identities' names and the blocking of images of identities together. Therefore, in the present experiment, in order to remove those top-down binding cues, the identities' names were not presented on screen, and the images were not blocked by identity (see Figure 2.2.1). The order of images was randomised for each participant. The same attention task (are the eyes looking at the camera or not) was used as in Experiment 1; participants were on average 92.2% accurate.

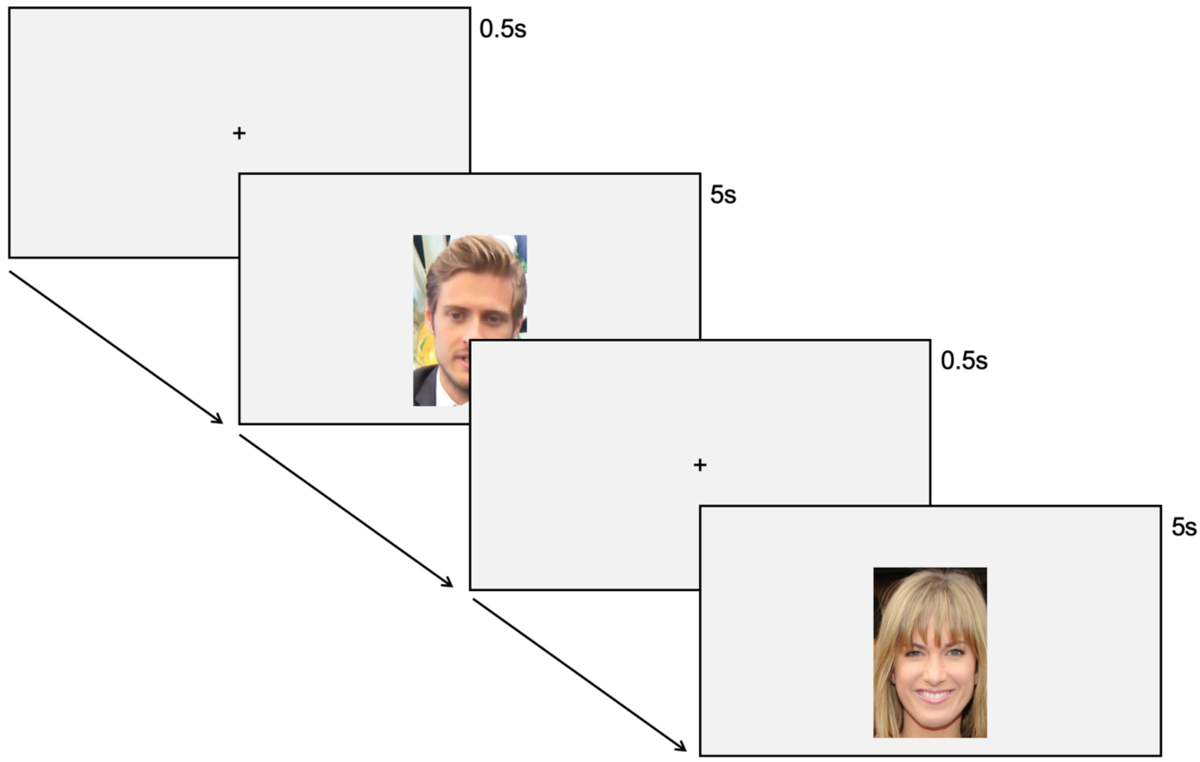


**Figure 2.2.1.** Example trials from the learning phase of Experiment 2. Images of each identity were shown in a random order with their names removed.

For the matching task, the design and procedure were the same as in the previous experiment. 2 participants failed the attention check and were removed from subsequent analyses. Of the remaining participants, they were on average 98.6% accurate. On average, the experiment lasted approximately 31 minutes.

## Results

*Match trials: Accuracy*

Mean accuracy for match trials is shown in Figure 2.2.2. A one way repeated measures ANOVA (factor: familiarity, levels: novel, low variation, high variation) was performed on the match accuracy data. The data did not violate the assumption of sphericity ($p$ = .662). A significant main effect of familiarity was found (F(2, 106) = 25.56, $p$ < .001, partial $\eta^2$ = 0.33). Bonferroni corrected pairwise comparisons revealed that identities learnt from low variation photos ($p$ < .001), or high variation photos ($p$ < .001) were matched significantly more accurately than novel identities. There was no significant difference in matching accuracy for low or high variation ($p$ = .591).

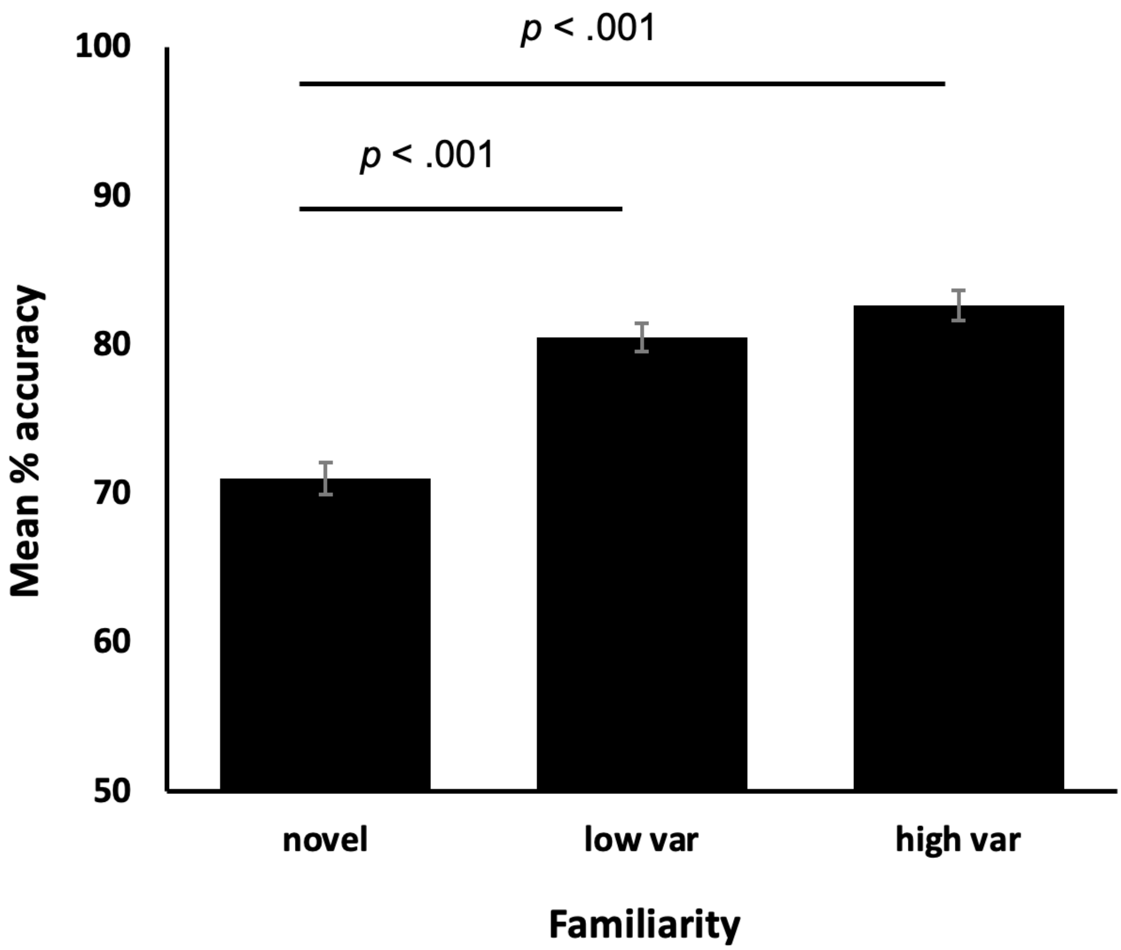



**Figure 2.2.2.** Mean accuracy for match trials for novel identities, identities learnt from low variation images and identities learnt from high variation images without top-down information.

*Match trials: RTs*

Mean RTs for match trials is shown in Figure 2.2.3. A one way repeated measures ANOVA was performed on the match trials RT data. The data did not violate the assumption of sphericity ($p$ = .255). A significant main effect of familiarity was found (F(2, 106) = 20.77, $p$ < .001, partial $\eta^2$ = 0.28). Bonferroni corrected pairwise comparisons revealed that identities learnt from high variation photos ($p$ < .001), or low variation photos ($p$ < .001) were correctly matched faster compared to novel identity matches. There was no significant difference in the reaction times between high variation and low variation identities ($p$ > .999).

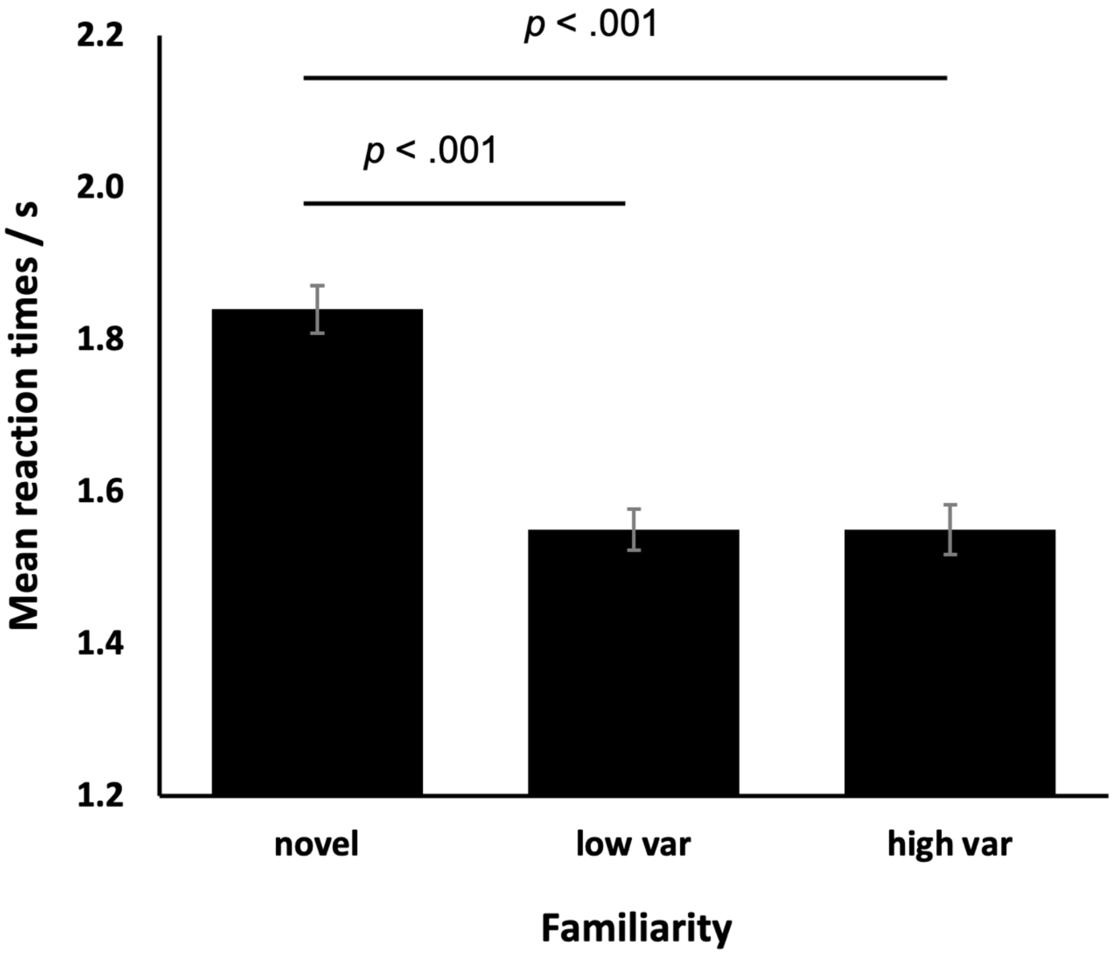



**Figure 2.2.3.** Mean reaction times for match trials for novel identities, identities learnt from low variation images and identities learnt from high variation images without top-down information.

*Mismatch trials: Accuracy*

Mean accuracy for match trials is shown in Figure 2.2.4. A one way repeated measures ANOVA was performed on the mismatch trials accuracy data. The data violated the assumption of sphericity (*p* = .048) therefore a Greenhouse-Geisser correction was applied. No significant main effects of familiarity were found ($F(1.80, 95.46) = 1.86$, *p* = .165, partial $\eta^2 = 0.03$).

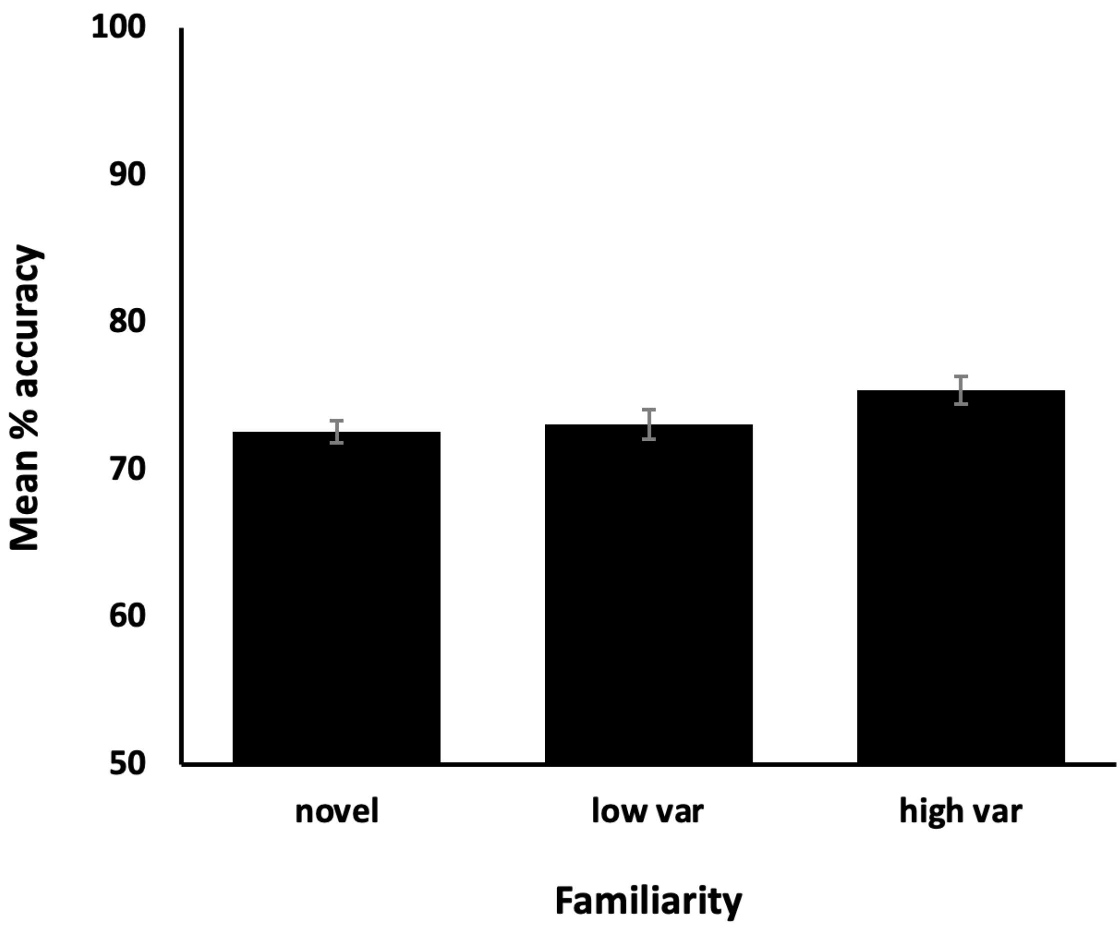



**Figure 2.2.4.** Mean accuracy for mismatch trials for novel identities, identities learnt from low variation images and identities learnt from high variation images without top-down information.

*Mismatch trials: RTs*

Mean RTs for mismatch trials is shown in Figure 2.2.5. A one way repeated measures ANOVA was performed on the mismatch trials RT data. The data violated the assumption of sphericity ($p < .001$) therefore a Greenhouse-Geiser correction was applied. A significant main effect of familiarity was found ($F(1.54, 81.44) = 4.04$, $p = .031$, partial $\eta^2 = 0.07$). There were no significant differences in the reaction times between novel and low variation identities ($p > .999$), novel and high variation identities ($p = .086$), nor between low and high variation identities ($p = .133$).

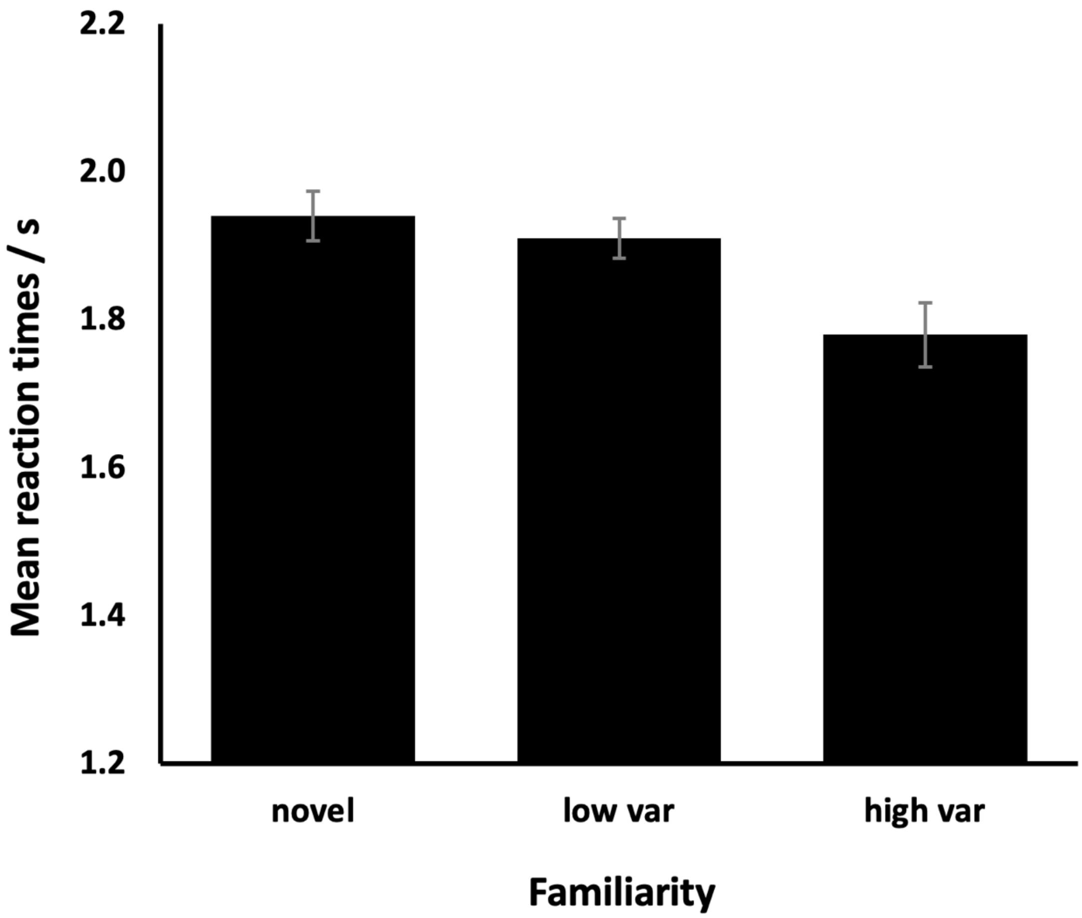



**Figure 2.2.5.** Mean reaction times for mismatch trials for novel identities, identities learnt from low variation images and identities learnt from high variation images without top-down information.

**Discussion**

In Experiment 2, when identities were learnt without top-down binding cues during the face learning phase, there was still no significant advantage of learning to recognise identities from high variation images over low variation images. The difference between identities learnt from low or high variation identities was not significantly different in either the accuracy or RT data. It may have been that as participants did not have information about which highly variable images to cohere together, they were less able to bind all of the images into a face representation, and thus performed slower and less accurately in the matching task, bringing performance more in line with identities learnt from low variation images.

However, it is hard to claim that the lack of top-down information in this experiment caused the accuracy and RTs of the high variation identities to become similar to that of the low variation identities, as the differences were not significant in Experiment 1. This experiment sought to find if there was a null effect in the context of finding a significant result in Experiment 1, however only a trend was found in Experiment 1. While it cannot be concluded from this experiment alone that removing top-down information removed the advantage usually found for viewing high variation images, it also cannot be concluded that top-down cues play no role in face learning. This is because there were areas for improvement in these first two experiments. For example, there may not have been a big enough difference between the low and high variation stimuli to find an effect. As the low variation stimuli were stills taken from videos, these contained lots of variation in facial expressions and speech as they were not posed photographs. High variation images however were photographs taken from internet searches, and as noted in previous research, such photos tend to be front facing smiling photographs (Jenkins et al., 2011).

Another possible issue with the first two studies is that multiple factors were varied at the same time. By aiming to test learning with top-down information in Experiment 1 by blocking identities and presenting a name, and testing learning without top-down cues in Experiment 2 by unblocking the identities and removing the name, two factors were manipulated at the same time: the presence/absence of top-down information, and blocked/unblocked experiment design. As two things were manipulated at the same time, this may have had an effect on the results. Therefore, the following two experiments aim to again test if top-down cues play a role in a face learning, and several key changes were made to hold the experiment design constant, increase the difference between the low and high variation stimuli, and multiple small improvements to increase power.

# Experiment 3: Face learning with top-down information

## Introduction

The previous 2 experiments lend some support to the idea that top-down information helps to cohere together variables instances of a face to form a face representation. In Experiment 1, when identities were learnt with top-down information, identities learnt from high variation images were matched slightly more accurately and faster than identities learnt from low variation images, however, the results only trended in the expected direction but did not achieve significance. In Experiment 2, when the top-down information was removed from the learning process, there was no significant difference between identities learnt from low or high variation images. However, without a significant result in Experiment 1, the results of Experiment 2 are less compelling.

Therefore, in the following 2 experiments, an improved attempt was made at finding this pattern. Multiple changes were made: the number of identities that participants learnt to recognise was increased from 10 to 16, therefore the number of identities learnt from low/high variation was increased from 5 to 8. Furthermore, the collected sample size was increased from 60 (54 final participants in Experiment 1) to 70 participants (65 final participants in Experiment 3). Therefore, the number of observations within each cell of the design increased from 2160 trials in Experiment 1 to 4224 trials in Experiment 3, almost double.

Changes were also made to the stimuli: for the low variation stimuli, new stills were selected which were less variable than the low variation images in Experiments

1 and 2. Also, some of the images were substituted for new higher quality images, and some were cropped more closely around the face. Furthermore, some of the celebrity names were Anglicised which would be easier for British participants to read, and they would therefore spend less time looking at the names and more time looking at the faces. During the learning phase, feedback was added to the attention task – “response recorded” appeared for 1s when participants pressed a key to indicate if the eyes were looking at the camera or not, to increase task engagement and attention.

Another key experiment change was the design. As discussed above, Experiment 1 employed a block design and Experiment 2 was unblocked. Therefore, in the following 2 experiments, both have the same unblocked design, so that only the variable of interest (the presence of top-down information) was manipulated. Finally, the inclusion criteria were modified and more were included. Rather than participants having to score 80% or above in the matching task attention check, in the following two experiments this was increased to 85%. Also, further inclusion criteria were added: participants had to complete at least 90% of trials and achieve at least 80% accuracy to be included in the data analysis.

Experiment 3 aimed to replicate the effect that learning a face from highly variable images benefits subsequent recognition. It was hypothesized that identities learnt from highly variables images would be matched more accurately than identities learnt from low variation images.

**Methods**

The following two experiments were pre-registered on the Open Science Framework (OSF). Full details of the hypotheses, experiment design, data exclusion criteria, analysis plan, stimuli, experimental scripts and raw data and can be found at https://osf.io/upj9g/.

***Participants***

66 right-handed participants (52 females, 14 males, mean age = 30.89, SD = 6.83) took part in this experiment. 4 additional participants were excluded from the analysis (3 participants experienced data recording issues and 1 participant got the response keys the wrong way round). Participants were recruited from Prolific and were paid £5 to compensate their time. All participants self-reported as being British, currently residing in the United Kingdom, had normal colour-vision, and normal or corrected-to-normal vision. Informed consent was gained, and the experiment was approved by the University of York Psychology Ethics Board.

***Materials***

Images of 6 new identities were collected for use in this experiment, in addition to the 10 identities from Experiments 1 and 2, totalling 16 foreign celebrities (8 female). For 2 of the original 10 identities, of the high variability learning images, 1 image each was swapped out for a new higher quality image. For the low variability learning images, new images were acquired for the original 10 identities. For 3 of those identities, stills were taken from new video clips, and for the remaining 7 identities the same video clips were used but new stills were selected. The new stills were chosen to display less within-person variation than the stills from Experiments 1 and 2, and the new videos for 3 of the identities were chosen for improved image quality.

For the images used in the matching task, 1 identity had 4 new images and another identity had 3 new images selected, which were higher quality and bore a better resemblance. A new foil was selected for one of the identities which bore better resemblance to them. For 2 of the foil identities, the same images were used but the images were cropped closer around the face. For the novel identities, the same images of the 40 identities and foils from Experiments 1 and 2 were used. Images of 24 new identities and 24 new foils were also collected, 3 images of each novel identity and 1 image of each foil, totalling 64 novel identities and 64 novel foils. See Appendix B for further stimuli examples.

***Design and Procedure***

*Learning Phase*

In order to keep the experiment design the same across Experiments 3 and 4, participants saw 10 photographs of each of 16 unfamiliar foreign celebrities, intermixed in a random order. This time, 8 of the identities were represented by high variation images and 8 represented by low variation images, counterbalanced across participants, producing 160 trials in total. At the start of a trial, a fixation cross appeared in the centre of the screen for 0.5s on a grey background. A face was then displayed centrally for 5s on a coloured background and their name displayed at the top of the screen. Each identity was paired with unique 'context' information, a distinct background-colour and a name (see Figure 2.3.1). Real names of the foreign celebrities were used. In an effort to improve the design of Experiment 1 and 2, the names were Anglicised for some of the identities so participants did not spend a long time looking at the names.

Participants again performed the attention check during the learning phase. They judged whether the eyes were looking at the camera or not by pressing 1 or 0; buttons were counterbalanced across participants. Again, in an effort to improve upon the experiment design, the message "response recorded" was displayed for 1s when participants pressed a key in order to maintain task engagement. Participants had to complete at least 90% of trials and achieve at least 80% accuracy to be included in the data analysis. On average, the participants responded to 99.73% of trials and were 94.84% accurate. The learning phase lasted 14 minutes and 40 seconds.

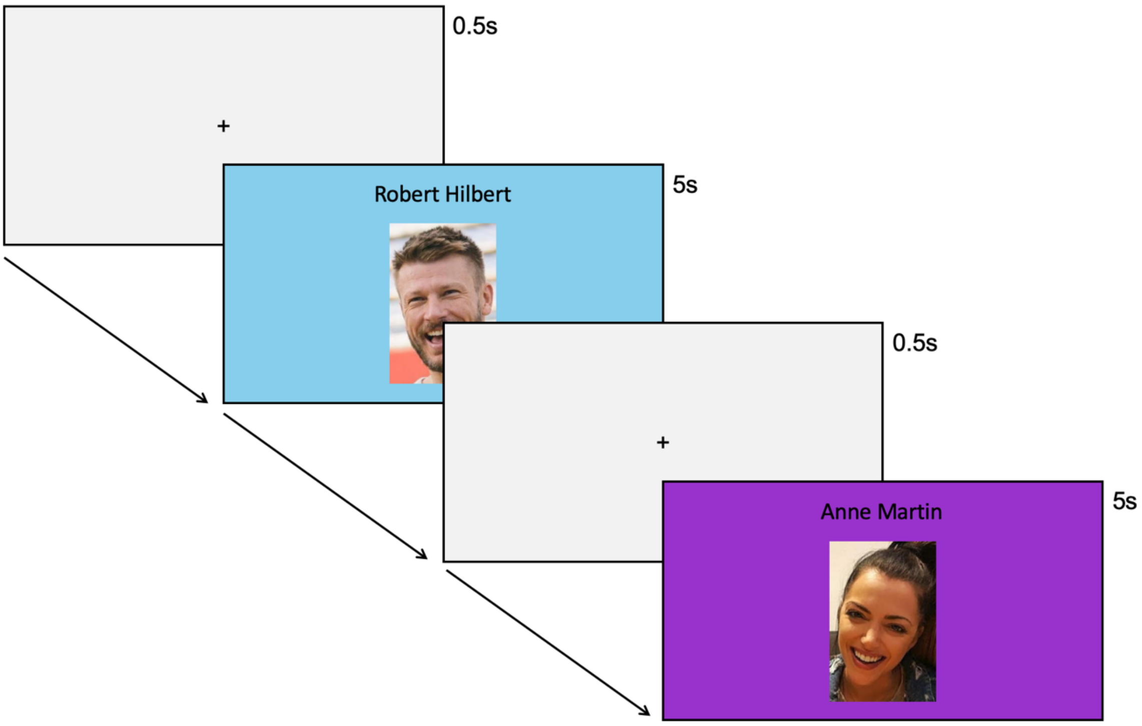



**Figure 2.3.1.** Example trials from the learning phase of Experiment 3. Images of each identity were paired with a unique name and background colour and shown in a random order.

*Matching Task*

The matching task was identical to that of Experiment 1 and 2. Participants indicated if pairs of photographs displayed side-by-side depicted the same person or different people; buttons were counterbalanced across participants. Again, trials began with a fixation cross presented in the centre of the screen for 0.5s followed by the pair of faces. The images remained on screen until participants made a response. Half of the trials were match trials and half were mismatch; there were 8 match and 8 mismatch trials for each learned identity (128 trials for low variation identities and 128 trials for high variation identities in total), and 1 match and 1 mismatch trial for each of the 64 novel identities (128 trials). The order of trials was randomised for each participant. In the mismatch trials, the presentation of the foil ID was presented on the left or right equally often. Participants were asked to respond as quickly and accurately as possible. To ensure participants paid attention, the same 20 identical image pairs were included (10 male, 10 female) where participants would be expected to display ceiling effects. Participants had to achieve at least 85% accuracy to be included in the data analysis. Participants were 98.56% accurate on these trials. The matching task lasted on average 29 minutes.

This experiment had 2 independent variables, one with 2 levels and one with 3 levels, producing 6 conditions. The familiarity level was either novel, low variation identity, or high variation identity. The trial type was either match or mismatch. Matching accuracy and RTs were recorded. On average, the experiment lasted approximately 44 minutes.

**Results**

Match and mismatch trials were analysed separately as in previous studies (White et al., 2014; Ritchie & Burton, 2017).

*Match trials: Accuracy*

Match trials of faces learnt from high variability photographs produced the highest accuracy, followed by low variation IDs, followed by novel IDs, as shown in Figure 2.3.2. A one way repeated measures ANOVA (factor: familiarity, levels: novel, low variation, high variation) was performed on the match accuracy data. The data violated the assumption of sphericity ($p$ = .002), so a Greenhouse-Geisser correction was applied. A significant main effect of familiarity was found ($F(1.70, 110.21) = 66.81$, $p < .001$, partial $\eta^2 = 0.51$). Bonferroni corrected pairwise comparisons revealed that IDs learnt from low variation ($p < .001$) and high variation ($p < .001$) images were matched significantly more accurately than novel IDs. IDs learnt from high variation images were also matched significantly more accurately than low variation IDs ($p < .001$).

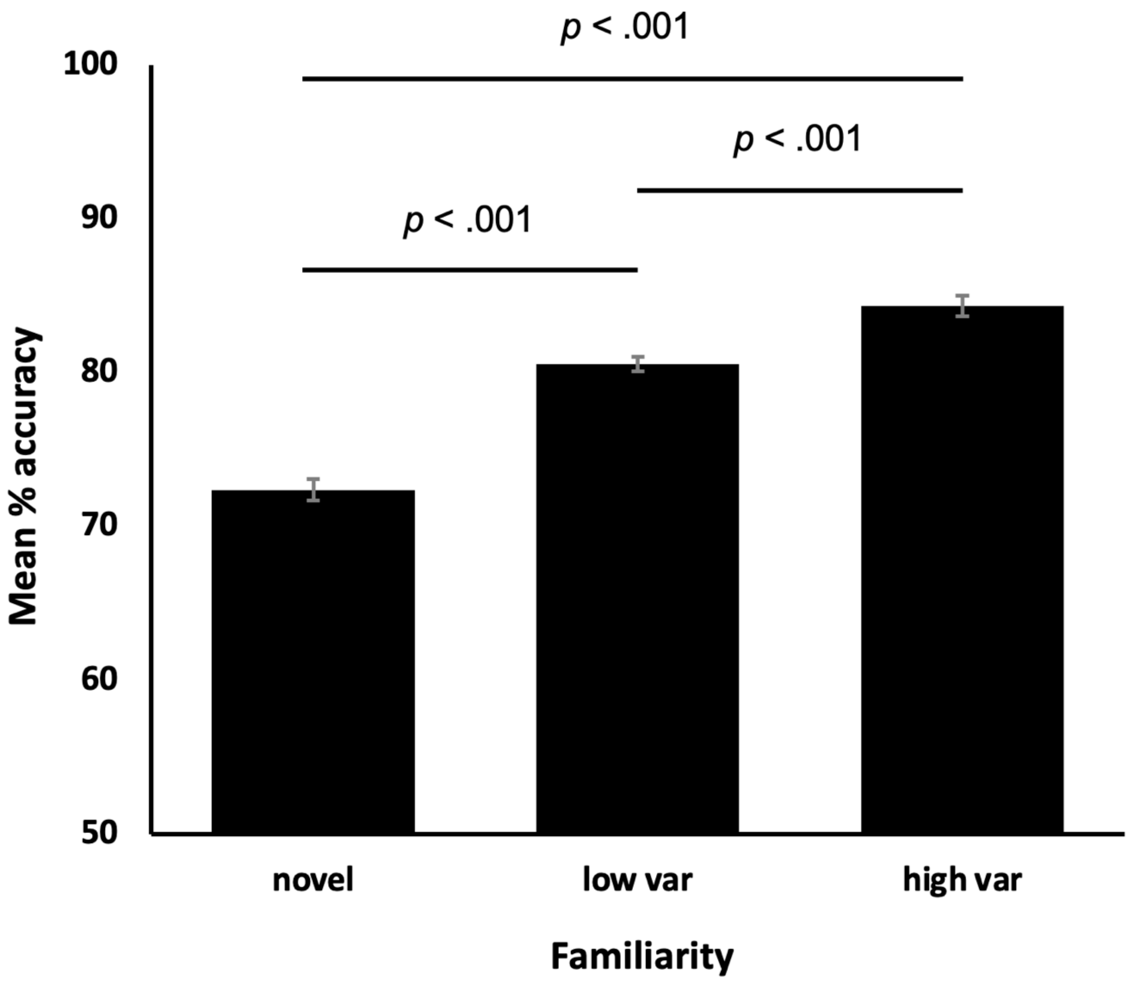


**Figure 2.3.2.** Mean accuracy for match trials for novel identities, identities learnt from low variation images and identities learnt from high variation images with top-down information.

*Match trials: RTs*

Faces learnt from high variability photographs were matched the fastest, followed by low variation IDs, followed by novel IDs, as shown in Figure 2.3.3. A one way repeated measures ANOVA was performed on the match trials RT data. Mauchly's test indicated the data violated the assumption of sphericity ($p$ < .001), therefore a Greenhouse-Geisser correction was applied. A significant main effect of familiarity was found ($F(1.31, 85.06) = 43.59$, $p$ < .001, partial $\eta^2 = 0.40$). Bonferroni corrected pairwise comparisons revealed that IDs learnt from low variation ($p$ < .001) and high variation ($p$ < .001) images were matched significantly faster than novel IDs. IDs learnt from high variation images were also matched significantly faster than low variation IDs ($p$ < .001).

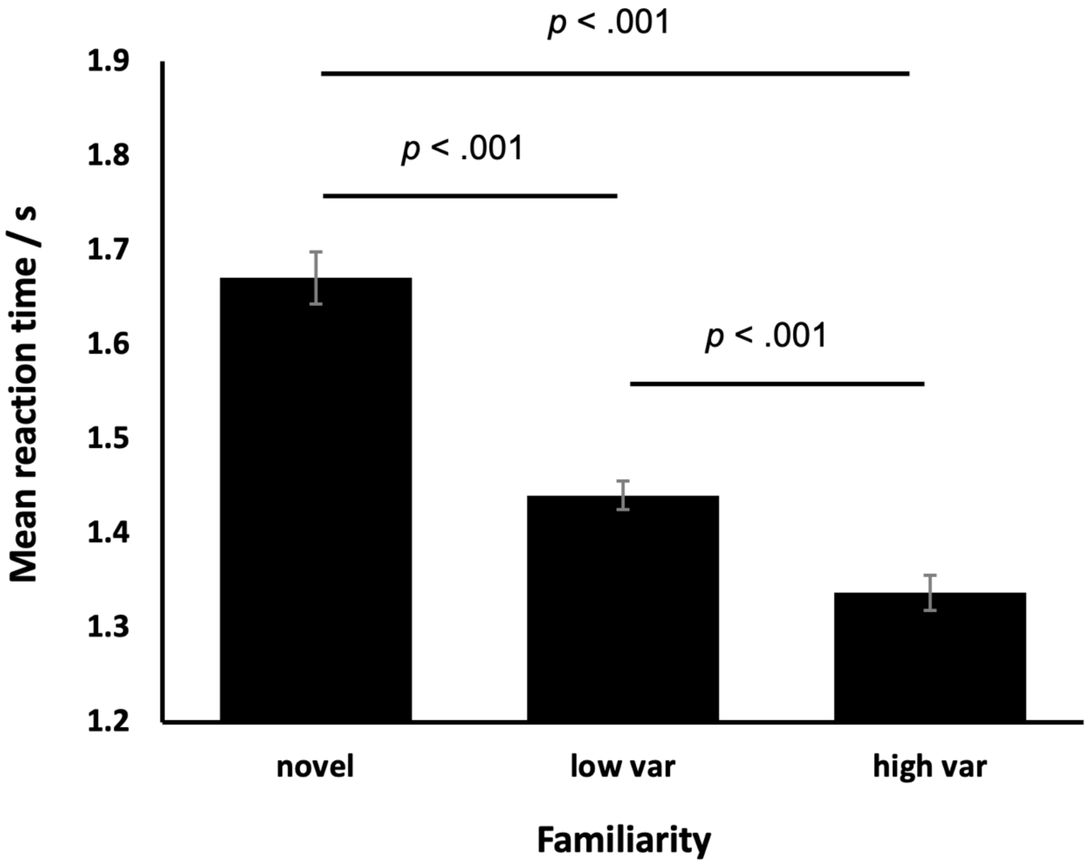



**Figure 2.3.3.** Mean reaction times for match trials for novel identities, identities learnt from low variation images and identities learnt from high variation images with top-down information.

*Mismatch trials: Accuracy*

Mismatch trials of novel faces produced the highest accuracy, as shown in Figure 2.3.4. A one way repeated measures ANOVA (factor: familiarity, levels: novel, low variation, high variation) was performed on the mismatch accuracy data. The data violated the assumption of sphericity ($p$ = .017), so a Greenhouse-Geisser correction was applied. A significant main effect of familiarity was found ($F(1.79, 116.08) = 4.38$, $p$ = .018, partial $\eta^2$ = 0.06). Bonferroni corrected pairwise comparisons revealed that novel IDs were matched significantly more accurately than IDs learnt from low variation ($p$ = .010) and high variation ($p$ = .036) images. The difference between IDs learnt from high variation and low variation images was not significant ($p$ > .999).

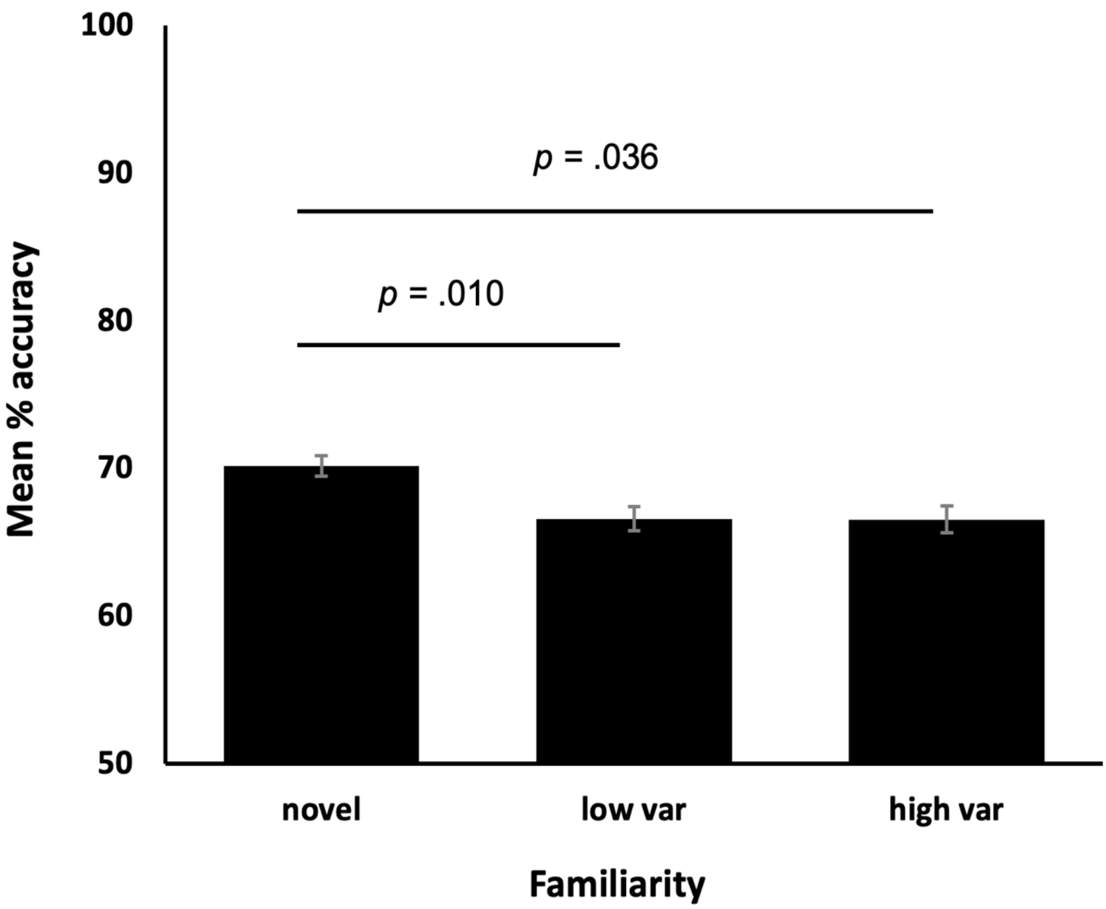



**Figure 2.3.4.** Mean accuracy for mismatch trials for novel identities, identities learnt from low variation images and identities learnt from high variation images with top-down information.

*Mismatch trials: RTs*

Figure 2.3.5 shows the mean reaction times across conditions for mismatch trials. A one way repeated measures ANOVA was performed on the mismatch trials

RT data. Mauchly's test indicated the data violated the assumption of sphericity ($p$ = .007), therefore a Greenhouse-Geisser correction was applied. No significant effect of familiarity was found ($F(1.75, 113.50) = 4.17$, $p$ = .022, partial $\eta^2 = 0.06$). Bonferroni corrected pairwise comparisons revealed that IDs learnt from low variation images were matched significantly faster than novel IDs ($p$ = .023). There were no significant differences between novel identities and high variation identities ($p$ = .280) or between low variation and high variation identities ($p$ = .814).

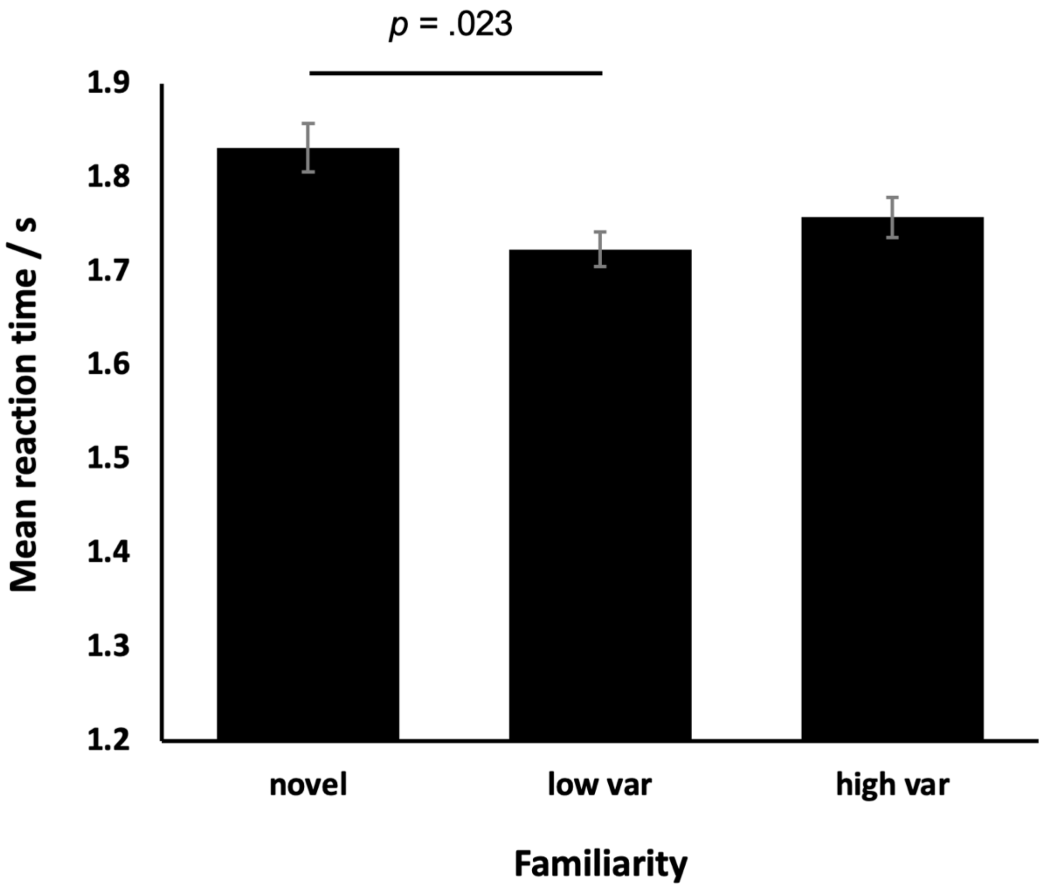



**Figure 2.3.5.** Mean reaction times for mismatch trials for novel identities, identities learnt from low variation images and identities learnt from high variation images with top-down information.

**Discussion**

In this experiment, the pattern of results found in Ritchie and Burton (2017) was successfully replicated. In both the accuracy and RT data, there was a benefit to learning identities from highly variables images compared to low variation images. This suggests that participants were able to able to cohere together the highly variables images into a face representation, which lead to better performance in the match trials of the subsequent matching task.

The mismatch trials produced an unexpected result. Novel identities were matched more accurately than both the identities learnt from low and high variation images, but they were matched more slowly than identities learnt from low variation images, and trended towards being matched more slowly than identities learnt from high variation images. This could suggest a speed-accuracy trade off, participants were more accurate with the novel identities however they were slower. As this result was not found in Ritchie and Burton (2017) or Experiments 1 and 2 here, it could perhaps be due to the different experimental design used. In Ritchie and Burton (2017) and Experiments 1 and 2 here, images of identities were blocked together, however, in the present experiment, the identities were unblocked. It could be that viewing long unblocked sequences of face images lead to an improvement on telling identities together but a cost on telling people apart.

The aim of this experiment was to make an improved attempt on Experiment 1 and find replicate the finding that viewing high variation images improves face learning. This was successfully done and the trend found in Experiment 1 became significant differences in the present experiment. Therefore, the following experiment aimed to test if removing top-down information would produce a null-effect where a significant one was found here. As in Experiment 2, the same images were used in Experiment

4 as in Experiment 3, only top-down information was removed, the background colour and names were removed. It is predicted that participants will not be able to cohere together the highly variable images as accurately without these top-down cues, and therefore participants will perform similarly with identities learnt from low or high variability images.

# Experiment 4: Face learning without top-down information

## Introduction

In the previous experiment, a significant benefit of learning identities from high variation images was found compared to low variation images, when paired with top-down information. The following experiment aims to test if top-down information is required for face learning by removing it. If top-down information is necessary to bind together variable instances of a face into a representation, then it is hypothesized that in this experiment when such information is removed, that identities learnt from high variation images will not benefit from the high variation and will be matched with similar accuracy and speed to identities learnt from low variation images.

## Methods

### *Participants*

65 new right-handed participants (51 females, 14 males, mean age = 28.71, SD = 6.22) took part in this experiment. 5 additional participants were excluded from the analysis (1 participant failed to meet the learning phase accuracy criterion, 2 participants failed to meet the matching task attention check accuracy criterion, and 2 participants experienced data recording errors). Participants were again recruited from Prolific and were paid £5 to compensate their time. All participants self-reported as being British, currently residing in the United Kingdom, had normal colour-vision, and

normal or corrected-to-normal vision. Informed consent was gained, and the experiment was approved by the University of York Psychology Ethics Board.

***Materials, Design and Procedure***

Experiment 4 was identical to Experiment 3 in all ways apart from the key manipulation – top-down information was removed. In the learning phase, the images of IDs were all presented on grey backgrounds and names were no longer presented above the images (see Figure 2.4.1). The experiment design, however, remained the same, the trials were again shown in a randomised order.

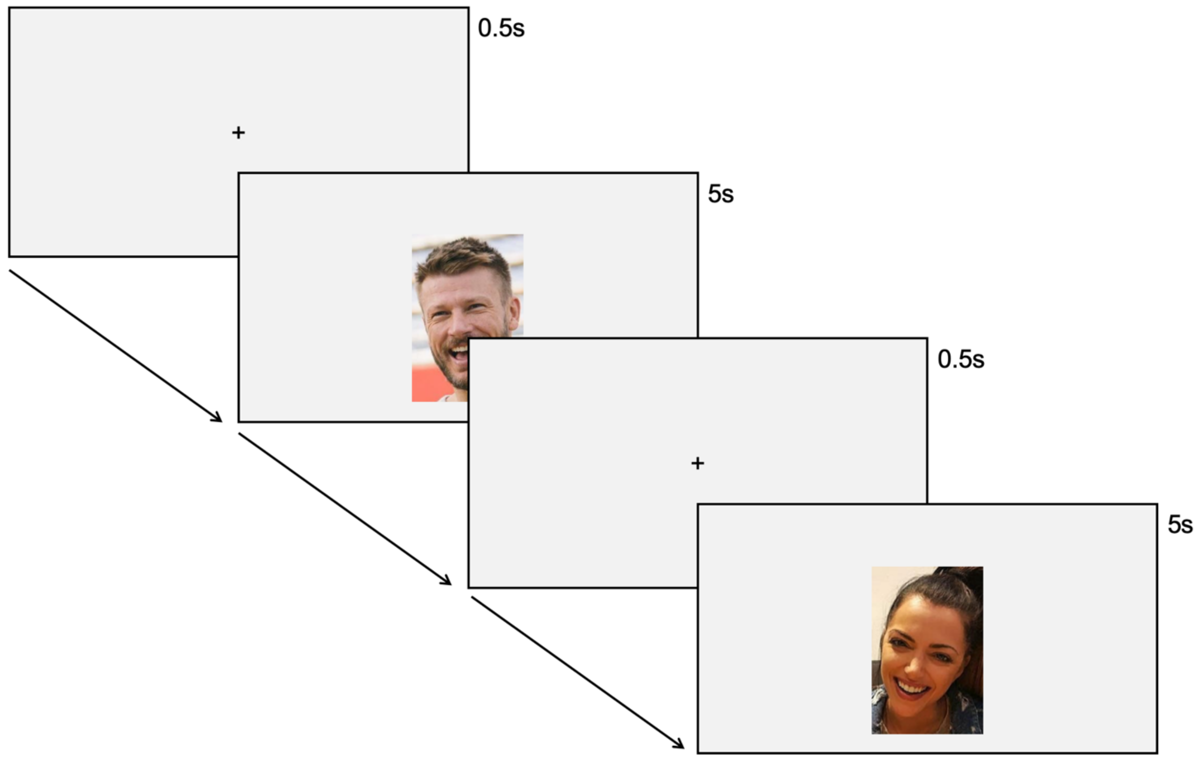



**Figure 2.4.1.** Example trials from the learning phase of Experiment 4. Images were presented on grey backgrounds without their name presented above and shown in a random order.

1 participant failed to meet the learning phase accuracy criterion and 2 participants failed to meet the matching task attention check accuracy criterion. On average, the remaining participants responded to 99.61% of learning trials, were 94.05% accurate, and were 98.31% accurate on the matching task attention check. On average, the experiment lasted approximately 43 minutes.

**Results**

Match and mismatch trials were analysed separately as in Experiment 3.

*Match trials: Accuracy*

Figure 2.4.2 shows the mean performance across conditions for match trials. A one way repeated measures ANOVA (factor: familiarity, levels: novel, low variation, high variation) was performed on the match accuracy data. The data violated the assumption of sphericity ($p < .001$), so a Greenhouse-Geisser correction was applied. A significant main effect of familiarity was found ($F(1.51, 96.75) = 63.75$, $p < .001$, partial $\eta^2 = 0.50$). Bonferroni corrected pairwise comparisons revealed that IDs learnt from low variation ($p < .001$) and high variation ($p < .001$) images were matched significantly more accurately than novel IDs. Crucially, the difference between IDs learnt from high variation and low variation images was not significant ($p = .066$).

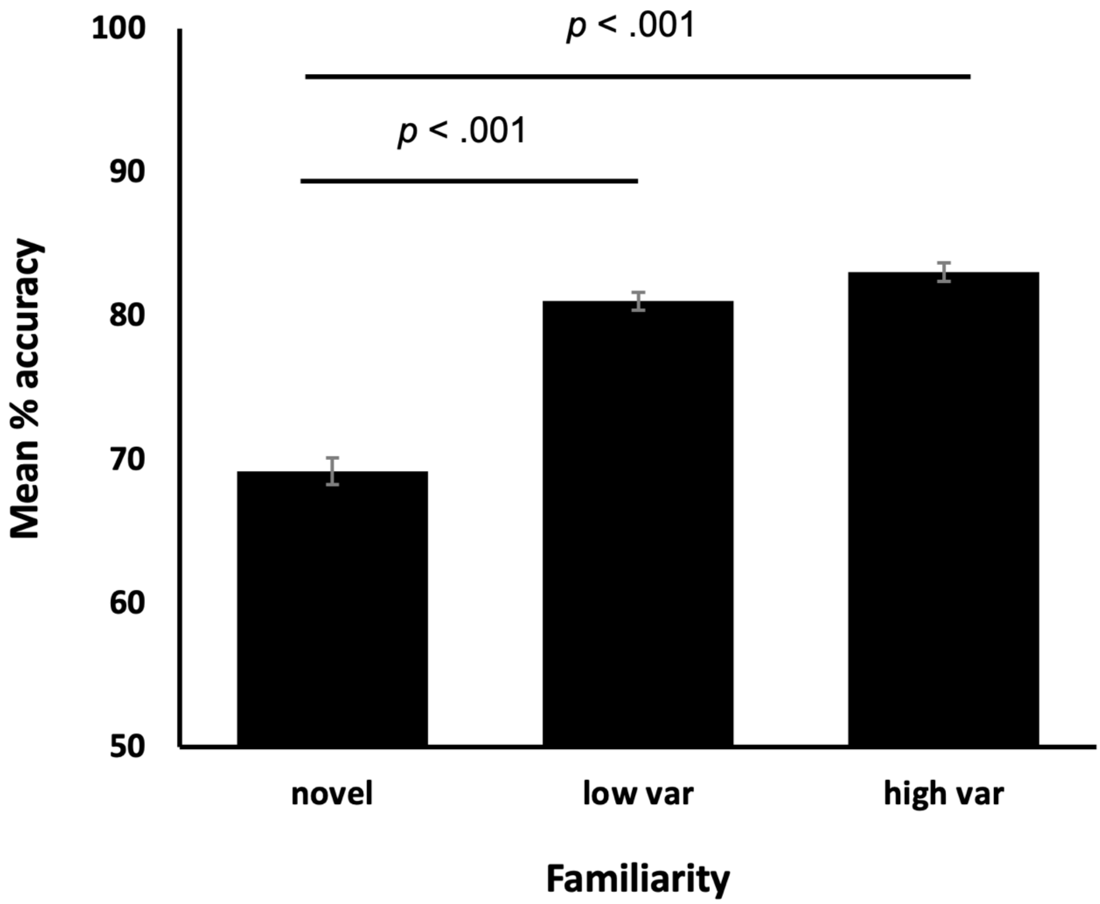



**Figure 2.4.2.** Mean accuracy for match trials for novel identities, identities learnt from low variation images and identities learnt from high variation images without top-down information.

*Match trials: RTs*

Figure 2.4.3 shows the mean reaction times across conditions for match trials. A one way repeated measures ANOVA was performed on the match trials RT data. Mauchly's test indicated the data violated the assumption of sphericity ($p$ < .001) so a Greenhouse-Geisser correction was applied. A significant main effect of familiarity was found (F(1.47, 94.06) = 53.37, $p$ < .001, partial $\eta^2$ = 0.45). Bonferroni corrected pairwise comparisons revealed that IDs learnt from low variation ($p$ < .001) and high variation ($p$ < .001) images were matched significantly faster than novel IDs. IDs learnt from high variation images were also matched significantly faster than low variation IDs ($p$ = .004).

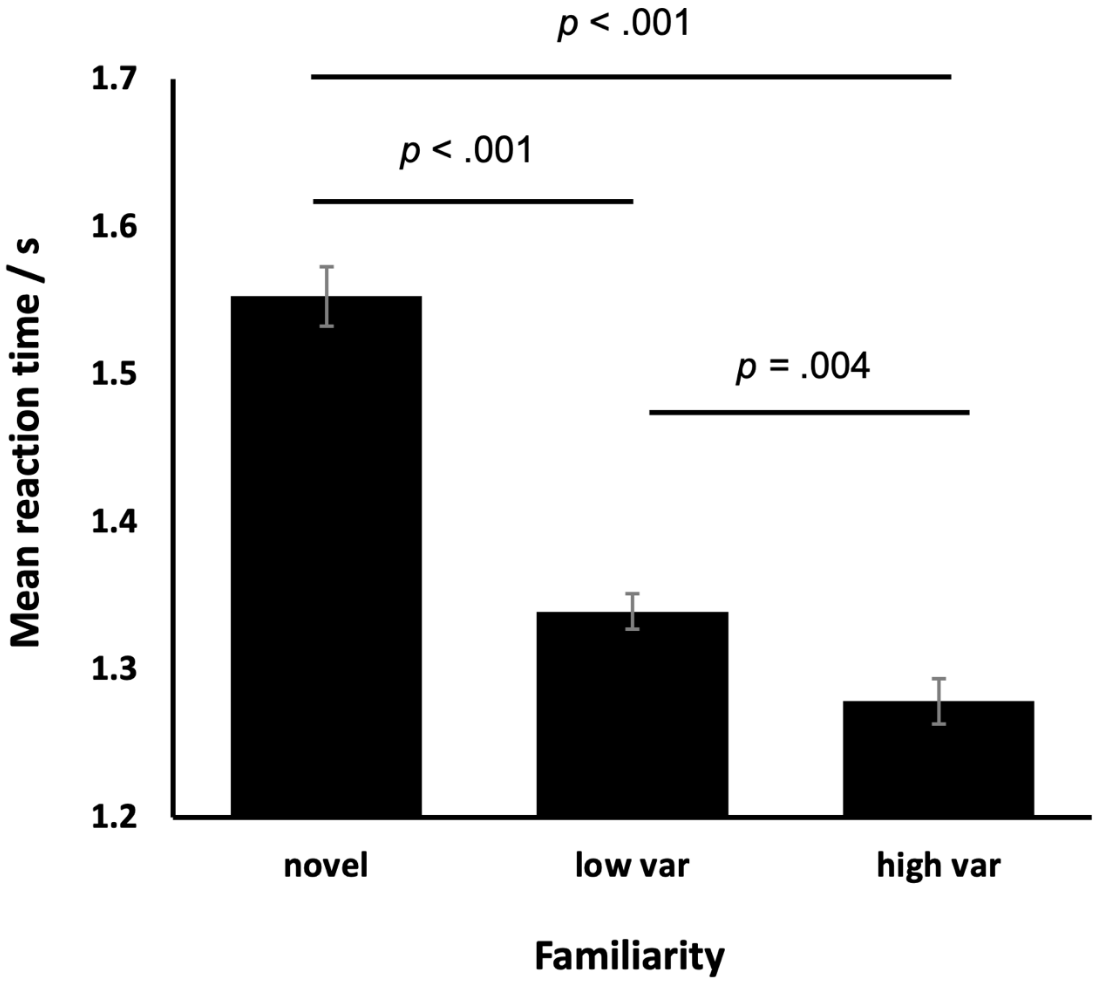



**Figure 2.4.3.** Mean reaction times for match trials for novel identities, identities learnt from low variation images and identities learnt from high variation images without top-down information.

*Mismatch trials: Accuracy*

Figure 2.4.4 shows the mean matching performance across conditions for mismatch trials. A one way repeated measures ANOVA (factor: familiarity, levels: novel, low variation, high variation) was performed on the mismatch accuracy data. The data violated the assumption of sphericity ($p$ = .044), so a Greenhouse-Geisser correction was applied. No significant main effect of familiarity was found (F(1.83, 116.97) = 0.88, $p$ = .409, partial $\eta^2$ = 0.01).

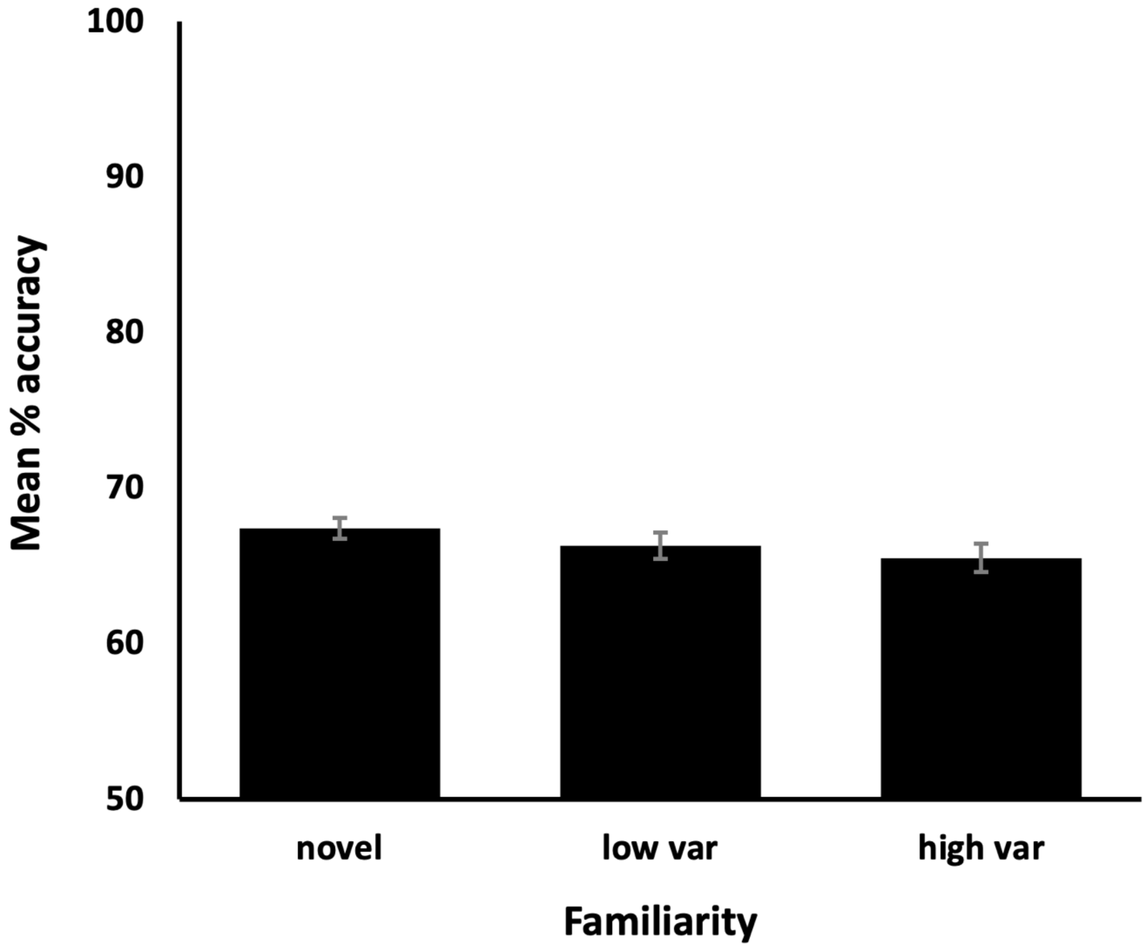



**Figure 2.4.4.** Mean accuracy for mismatch trials for novel identities, identities learnt from low variation images and identities learnt from high variation images without top-down information.

*Mismatch trials: RTs*

Figure 2.4.5 shows the mean reaction times across conditions for mismatch trials. A one way repeated measures ANOVA was performed on the mismatch trials RT data. Mauchly's test indicated the data did not violate the assumption of sphericity (*p* = .370). A significant main effect of familiarity was found (F(2, 128) = 3.14, *p* = .047, partial $\eta^2$ = 0.05). There were no significant differences in the reaction times between novel and low variation identities (*p* = .186), novel and high variation identities (*p* > .999), nor between low and high variation identities (*p* = .074).

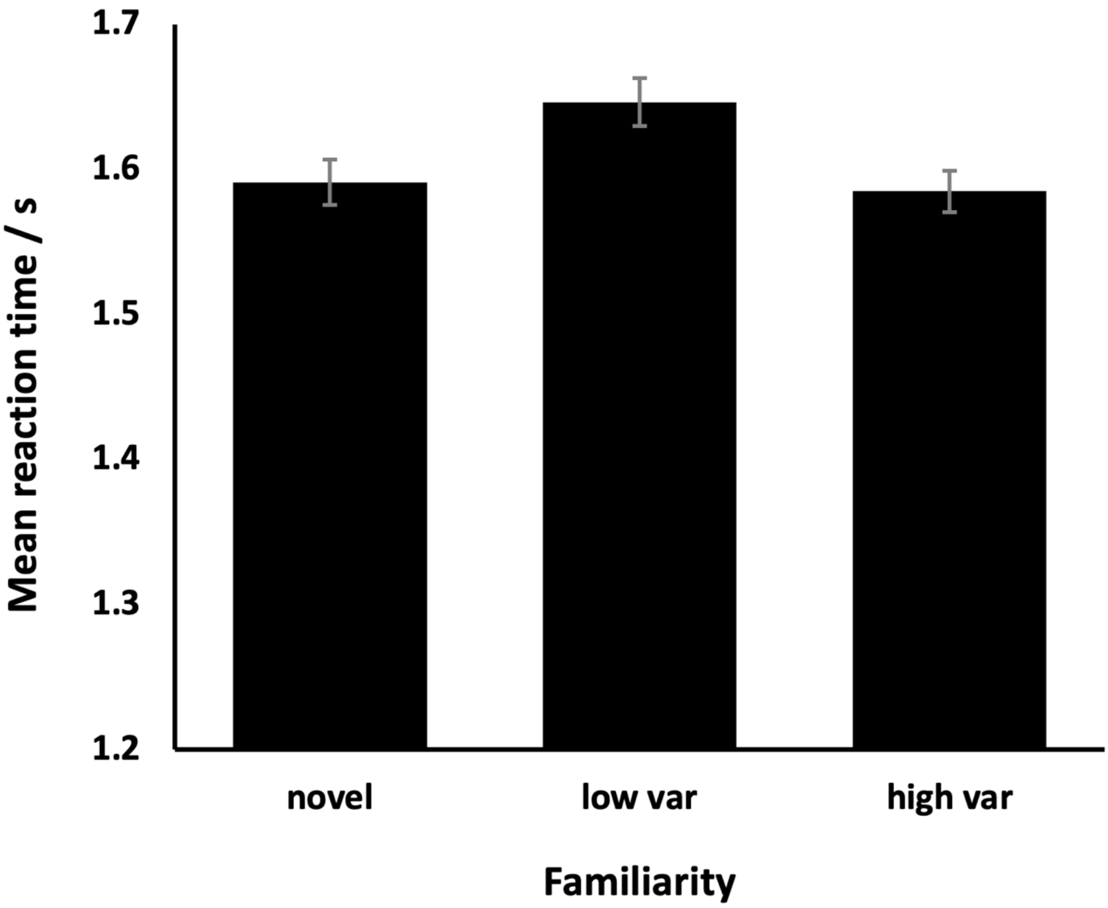



**Figure 2.4.5.** Mean reaction times for mismatch trials for novel identities, identities learnt from low variation images and identities learnt from high variation images without top-down information.

**Further Analysis**

Experiments 3 and 4 showed different patterns, Experiment 3 was modulated by the amount of variation but Experiment 4 was not. Ideally a combined within-subjects ANOVA would be conducted to demonstrate this difference. However, Experiments 3 and 4 were pre-registered and used different participants, therefore a combined mixed ANOVA will be comparatively statistically weak. The exploratory analysis is presented below.

*Match trials: Accuracy*

A 2x3 mixed ANOVA (factor: experiment, levels: Experiment 3, Experiment 4; factor: familiarity, levels: novel, low variation, high variation) was performed on the match accuracy data. The data violated the assumption of sphericity ($p < .001$), so a Greenhouse-Geisser correction was applied. A significant main effect of familiarity was found ($F(1.64, 211.19) = 128.01$, $p < .001$, partial $\eta^2 = 0.50$). The main effect of experiment was not significant ($F(1,129) = 0.45$, $p = .504$, partial $\eta^2 = 0.00$) and the interaction was not significant ($F(1.64, 211.19) = 2.31$, $p = .112$, partial $\eta^2 = 0.02$).

*Match trials: RTs*

A 2x3 mixed ANOVA (factor: experiment, levels: Experiment 3, Experiment 4; factor: familiarity, levels: novel, low variation, high variation) was performed on the match reaction time data. The data violated the assumption of sphericity ($p < .001$), so a Greenhouse-Geisser correction was applied. A significant main effect of familiarity was found ($F(1.55, 200.41) = 85.60$, $p < .001$, partial $\eta^2 = 0.40$). The main effect of experiment was not significant ($F(1,129) = 1.80$, $p = .183$, partial $\eta^2 = 0.01$) and the interaction was not significant ($F(1.55, 200.41) = 1.44$, $p = .239$, partial $\eta^2 = 0.01$).

*Mismatch trials: Accuracy*

A 2x3 mixed ANOVA (factor: experiment, levels: Experiment 3, Experiment 4; factor: familiarity, levels: novel, low variation, high variation) was performed on the mismatch accuracy data. The data violated the assumption of sphericity ($p < .001$), so a Greenhouse-Geisser correction was applied. A significant main effect of familiarity was found ($F(1.81, 233.53) = 4.40$, $p = .016$, partial $\eta^2 = 0.03$). The main effect of experiment was not significant ($F(1,129) = 0.22$, $p = .639$, partial $\eta^2 = 0.00$) and the interaction was not significant ($F(1.81, 233.53) = 0.78$, $p = .448$, partial $\eta^2 = 0.01$).

*Mismatch trials: RTs*

A 2x3 mixed ANOVA (factor: experiment, levels: Experiment 3, Experiment 4; factor: familiarity, levels: novel, low variation, high variation) was performed on the match reaction time data. The data violated the assumption of sphericity ($p < .001$), so a Greenhouse-Geisser correction was applied. The main effect of familiarity was not significant ($F(1.68, 216.70) = 0.55$, $p = .545$, partial $\eta^2 = 0.00$). The main effect of experiment was not significant ($F(1,129) = 1.90$, $p = .171$, partial $\eta^2 = 0.01$) and the interaction was not significant ($F(1.68, 216.70) = 2.77$, $p = .074$, partial $\eta^2 = 0.02$).

**Discussion**

Matching trials accuracy showed an expected pattern of exposure, that learned identities were matched more accurately than novel identities, however, unlike Experiment 3, this was no longer modulated by the level of variation observed during learning. Participants were similarly accurate at matching identities learnt from low or high variation images, however, participants were significantly faster at matching identities learnt from high variation images compared to low variation images. This demonstrates that removing top-down cues from the face learning phase reduced the benefit of observing highly variable photographs, however, it did not abolish the effect completely.

In Experiment 3, when identities were learnt with top-down cues in the form of a name and coloured background, participants were significantly more accurate and faster at matching identities learnt from high variation images compared to low variation images. When the same experiment design was used and the same images, but top-down information was removed, participants were no longer more accurate at matching identities learnt from high variation images compared to low variation images. This suggests that participants were able to build face representations during the learning phase and cohere together the highly variable photographs in Experiment 3. In Experiment 4 however, as participants did not have the cue of the name and background to aid in image cohering, less images may have been successfully bound together, producing a representation containing less instances of the individual, producing poorer matching performance. However, as participants were significantly faster and non-significantly more accurate at matching high variation identities in Experiment 4 without top-down cues, this suggests top-down information is helpful in the face learning process but it not necessary.

# General Discussion

Taken together, these experiments suggest that top-down information about which images to cohere together is helpful but not necessary for face learning. In Experiment 1, when faces were learnt with top-down information, a trend was found that identities learnt from high variation images were more accurately and quickly matched than identities learnt from low variation images. In an improved attempt in Experiment 3, this same pattern was found but with significant differences. In Experiments 2 and 4, when top-down information was removed from the face learning process, the advantage gained from exposure to high variability images was reduced: identities learnt from low and high variability images were matched with similar accuracy.

The pattern found in Experiment 1 and the statistical outcomes of Experiment 3 fit with the findings of Ritchie and Burton (2017), that the more within-person variation observed during face learning, the better recognition accuracy. In Experiments 1 and 3, participants were able to use the broader range of within-person variation in the high variation condition to form generalisable face representations, which enabled them to recognise and match novel images of the identities in the matching task. For the low variability condition, as less within-person variation was seen during learning, performance was poorer in the matching task.

The advantage of viewing the high variability photos over the low variability photos in Experiments 1 and 3 may be due to the underlying face representations. Participants may have been able to use the top-down information to cohere together the highly variable images of identities into unique representations, to *tell them together*. Novel images of the faces were then recognised accurately in the matching task because the face representation contained enough within-person variation to be able to generalise and the novel instance was accurately matched to the face representation (Jenkins & Burton, 2011). These face representations can be thought of as sub-regions of face space (Valentine, 1991), a region within which any point represents an observed instance of that person. The centroid of the face region might be analogous to the average of a particular person. The advantage of this idea as face representations corresponding to a region of face space, rather than a specific point (Valentine, 1991) is that is allows for idiosyncratic within-person variation. One must learn how an individual varies and build up and shape that region of face space in order to be able to subsequently recognise them. This is consistent with previous work which has found that faces vary somewhat idiosyncratically (Burton et al., 2016; Lander & Chuang, 2005).

When this top-down information was removed in Experiments 2 and 4, participants were less able to tell images of the same identity together, relying instead on only bottom-up visual information from the images. Participants may not have been able to cohere all of the variable images into stable representations, perhaps perceiving different images of the same identity as depicting different people. This phenomenon is often found in card-sorting tasks where unfamiliar viewers, without any top-down information, tend to split photographs of single identities into multiple identity piles, perceiving there to be multiple different faces (Jenkins et al., 2011). However, this failure of *telling people together* can be alleviated when unfamiliar viewers are given top-down information, e.g. about the number of identities present (Andrews et al., 2015). The results of Experiments 2 and 4 suggest that as participants did not have any top-down information to bind highly variable images of the same identity together, they were not able to tell all of the images together and formed smaller less representative face representations, minimizing the advantage of seeing and cohering high variation images, which produced similar matching accuracy for both low and high variability identities. Participants may have been able to tell low variation images together as they inherently look very similar, however, as they contain less within-person variation, this produces smaller face representations which are less able to generalise to novel instances of the identity.

These results also fit with computation model results. Kramer et al. (2018) found that computational models built with PCA (e.g. bottom up information) and LDA (top-down information) were better able to recognise novel images of familiar identities compared to models just using PCA. The models used highly variable images, therefore these results are similar to the results of Experiments 1 and 3, that when face learning was accompanied by top-down information, this may have helped to cluster these images together for face representation building. Unlike in the computational models, low variability images did not benefit from top-down information, as mentioned above they are presumably easier to cohere together purely from bottom-up information as they look highly similar.

How might top-down information benefit face learning in day to day life? When encountering a new unfamiliar face in a conversation, variation in the appearance of the face is observed, from changing facial expressions, head angle, speech and so forth. Such changes in unfamiliar faces are known to disrupt recognition (Bruce, 1982; Hancock, Bruce & Burton, 2000). However, during a conversation we know we are speaking to the same person. We can then confidently cohere the variation seen into a single representation. If we then arrange to meet our new acquaintance for coffee, there is expectation-driven top-down support for recognising this person at the subsequent meeting, in addition to bottom-up visual cues. The variation added by this new encounter can be added to one's face representation, making it more generalisable, and so on for future meetings. It has recently demonstrated that such real-world encounters can result in robust behavioral face learning effects (Sliwinska et al., 2022).

These experiments aimed to test if top-down information is beneficial to face learning by focusing on the differences between identities learnt from low and high variability images. Whilst the difference between low and high variation identities learnt with top-down information in Experiment 3 was significantly different and the difference between low and high variation identities learnt without top-down information in Experiment 4 was not significantly different, this is not the biggest pattern found in these experiments. It seems that either learning regime, whether from low or high variation images, gives a larger benefit to learning, compared to the difference between low and high variation conditions. This is shown in the mixed ANOVA which found a significant main effect of familiarity but failed to find a significant main effect of experiment (presence/absence of top-down information) or an interaction. For example, in Experiment 1 when identities were learnt with top-down information, the performance difference between low and high variation IDs was just 1.85%, however the difference between low variation IDs compared to novel IDs was 6.53% and the difference between high variation IDs and novel IDs was 8.38%. The effect of any learning (e.g. low/high variation IDs compared to novel IDs) was again bigger than the effect of high variation over low variation IDs in Experiment 3: the performance difference between low and high variation IDs was 3.76%, however the difference between low variation IDs compared to novel IDs was 8.19% and the difference between high variation IDs and novel IDs was 11.96%.

Interestingly, this pattern was again found in Experiments 2 and 4 when top-down information was removed from the learning process: the difference between low and high variability IDs compared to novel IDs was bigger than the difference between low and high variability IDs, and there was a non-significant trend of high variation IDs being matched slightly more accurately than low variation IDs. In Experiment 2, the difference between low and high variation IDs compared to novel IDs was 9.49% and 11.62% respectively. Similarly, in Experiment 4, the difference between low and high variation IDs compared to novel IDs was 11.83% and 13.87% respectively.

This pattern suggests that top-down information is not essential for face learning, it is only beneficial. In Experiments 1 and 3 when top-down information was provided to participants, this benefitted the identities learnt from highly variable images, IDs learnt from highly variable images were matched more accurately than IDs learnt from low variation images. However, the larger effect was observed between learnt identities (whether from high or low variation images) compared to novel identities. This shows that in the low variation condition, even seeing a small amount of variation benefitted participants significantly. They were able to cohere the low variation images together into a face representation, which enabled them to recognise novel images in the matching task. A larger benefit was seen for identities learnt from low variation images over novel identities, compared to the benefit seen for identities learnt from high variation images over low identities. As top-down information is perhaps not beneficial to low variation images, as they look inherently similar, and yet low variation identities were matched more accurately than novel identities, this shows that top-down information is not essential for face learning. It has a smaller role of providing a small benefit to high variation images over low variation images.

This finding is interesting as it was also found in Experiments 2 and 4 when no top-down information was provided during face learning. Participants were still able to cohere both the low variation and high variation images together and produced more accurate matching compared to novel identities even in the absence of top-down cues. Furthermore, there was a trend in both experiments that identities learnt from high variation images were matched more accurately than identities learnt from low variation images. This again suggests that face learning is primarily driven by bottom-up visual information, and is benefitted by top-down information. These findings fit with the vast literature which shows that familiar identities (here low and high variation IDs) are recognised much more accurately than unfamiliar faces (here novel IDs). This experiment suggests that the acquisition of familiarity is an effective and rapid process.

This pattern of results fits with other work on top-down effects in face learning. For example, in the second experiment in Andrews et al. (2015), participants completed a card sorting task either with top-down information (two-sort), or without top-down information (free-sort). Participants were instructed to make identity piles, 1 pile of photos per identity. The free-sort group made a mean of 6.8 piles, they were unable to tell the variables images together, in comparison to the two-sort group who were instructed to make 2 piles. The two-sort group, free-sort group and a new group who had not completed the card sorting task (no-sort) then completed a face matching task, using identities from the card sorting task and novel identities not seen before. The no-sort group were 75% accurate with the card sorting IDs, the free-sort group were 81% accurate and the two-sort group were 86% accurate. These results are in line with experimental results found in this chapter, that top-down information conferred a benefit, participants in the two-sort condition performed better in the matching task compared to the free-sort group. Furthermore, participants who sorted the cards with no top-down information (free-sort group) performed better than the participants who had not completed a card sorting task and were completely unfamiliar with the identities. This is akin to the findings here that identities learnt from low variation images were matched better than novel identities. This is interesting as participants in the free-sort group made a mean of 6.8 piles in the card sorting task, which suggests the struggled to cohere together the within-person variation across images, yet their improved performance in the matching task compared to unfamiliar participants suggests that they were partly able to cohere some images together using only bottom-up visual information, in order to make a face representation to then match images against in the matching task.

How faces are initially learnt is still not well understood, however, the results of this chapter suggest that bottom-up visual cues drive face learning, and when the input is highly variable, top-down information aids face learning. When learning to recognise a face, a viewer observes within-person variation of the face and begins to cohere these images together into a face representation perhaps mainly by using bottom-up visual information. Top-down information helps to support further image cohering when the images are highly variable, supporting the building of a face representation which is then able to match novel images of the identity. Future research is needed to examine the kinds of top-down cues used in day to day face learning and other factors which may benefit face learning.

# Chapter 3 – Face Inversion

## Introduction

The face inversion effect (FIE) refers to the disproportionately larger drop in accuracy when recognising inverted faces (accuracy upright – accuracy inverted) compared to other mono-oriented stimuli (Yin, 1969). This is a very important effect in the literature for a number of reasons: a) it is thought to provide evidence that faces are processed configurally, along with the part-whole effect and the face composite effect; b) as larger inversion effects are found for faces compared to objects, it is thought to provide evidence that faces are processed in a unique way; c) and it is a robust effect, and has been replicated many times over (McKone & Yovel, 2009) across a range of subdisciplines and research areas such as developmental psychology, neuropsychology, memory and perception studies, emotion and identity recognition (Valentine, 1988). The following literature review and discussion will focus primarily on the face inversion effect in regard to identity recognition in neurotypical adults.

Yin (1969) suggested that faces may be disproportionately affected by inversion compared to other stimuli as there is a ‘special factor’ relating to the processing of faces which is inhibited by inversion. In the seminal paper, Yin (1969) used an old-new recognition task with black and white photos of cropped faces, silhouettes of aeroplanes, stick figures of men in motion and black and white photographs of houses. In Experiment 1, he found that when the inspection and test series were presented inverted, faces produced the highest number of errors. However, in Experiment 2, when the inspection series was presented upright and the test series inverted, men in motion stick figures produced the highest number of errors.

Overall, the paper claims faces are disproportionately affected by inversion due to a special type of processing unique to faces being disrupted by inversion.

Since his paper, many researchers have aimed to discover what type of processing might be specific to faces and cause the FIE. One account, the holistic theory, states that facial features are not processed independently but interdependently, and the face is perceived as a whole, and such processing is inhibited by inversion. Evidence for this account of the FIE comes from Tanaka and Farah (1993). In a learning phase, they paired names with dot drawings of faces presented upright or inverted. At test, participants discriminated between pairs of individual features or faces which differed by one feature. For upright faces, they found that participants were better at recognising features presented in the whole face compared to in isolation, suggesting that whole faces were processed holistically. However, when faces were inverted, there was no difference between the whole face and isolated features, suggesting that inversion had disrupted the holistic processing of the whole face, eliminating the advantage. A potential issue with this experiment is the stimuli used were drawings made of dots. Such stimuli do not represent normal visual experience with faces, and may not be generalisable to real life faces. Stimuli which are representative of normal visual input would not be images created by a small number of dots, but full colour photographs, which are intact, i.e. not having the background, body, or external facial features removed.

Other studies provide evidence for a configural account of the FIE. The configural processing theory suggests that the perception of the metric distances between key facial features is inhibited by inversion. Mondloch et al. (2002) showed participants greyscale images of faces wearing surgical caps with their shoulders covered by material. Pairs of upright or inverted images were presented sequentially which either differed in the spacing of the features or the actual features themselves, and participants had to indicate if the images were the same or different. They found large inversion effects for spatial changes but only small inversion effects for the featural changes, suggesting it is configural processing which is particularly inhibited by inversion.

A similar account for the face inversion effect, the expertise hypothesis, differs only from the configural processing theory in that it suggests the face inversion effect is not face specific, but similar decreases in recognition due to inversion can be found for other stimuli which one has expertise with. For example, Diamond and Carey (1986) tested dog experts and novices with upright and inverted human faces and dogs. In a memory task, participants had to indicate which image in a pair they had seen before. For images of human faces, both dog experts and novices showed large inversion effects, however, for images of dogs, only dog experts showed large inversion effects. The authors suggest this demonstrates that configural processing of objects of expertise is disrupted by inversion rather than faces specifically.

In Valentine's (1988) key review on face inversion, he puts forth the position that faces are not special, i.e. that they are not processed in a unique way, and lends support to Diamond and Carey's (1986) expertise account. He does not agree with the strong view of Diamond and Carey that face inversion causes a change in processing strategy, from configural processing when upright to featural processing when inverted, however, his review supports a quantitative difference in processing between upright and inverted faces. His review covers literature which demonstrates that faces are processed configurally when upright, and this configural information is harder to perceive when inverted. For example, he describes the finding that as familiar faces are rotated away from upright, the reaction time to recognise them increased linearly.

One factor which might play a modulatory role in the face inversion effect is the familiarity of the face. Familiar and unfamiliar faces are processed differently, with familiar faces being recognised faster than unfamiliar faces (Clutterbuck & Johnston, 2002). Familiar face recognition is unharmed by changes in viewpoint and lighting (Hill & Bruce, 1996) and emotional expression (Bruce, 1982) unlike unfamiliar face recognition. Furthermore, familiar face recognition relies more on internal features as opposed to unfamiliar face recognition which relies more on external features (Ellis et al., 1979). Viewers can tolerate within-person variation of familiar faces, correctly grouping multiple photographs of a familiar identity together, whereas viewers cannot tolerate such variation in unfamiliar faces (Jenkins et al., 2011).

Taken together this evidence shows that familiar and unfamiliar faces are processed differently. However, mixed results have been found concerning the modulatory effect of familiarity on the face inversion effect. Megreya and Burton (2006) used greyscale images of unfamiliar and familiar cropped faces in a matching task, and found that upright unfamiliar face matching correlated with inverted but not upright familiar face matching. They also found that only familiar face matching suffered accuracy decreases when inverted, not unfamiliar face matching. Further evidence for larger inversion effects for familiar faces comes from Balas et al. (2010), who in an ERP study presented babies with upright or inverted images of their mothers or strangers. They found a main effect of orientation only in the mother’s face group, but not the stranger’s face group, again suggesting a modulatory role for face familiarity in the inversion effect. Furthermore, using a perceptual field paradigm, Wang et al. (2023) found large inversion effects for personally familiar faces, small inversion effects for moderately familiar faces and no inversion effect for unfamiliar faces. However, other work has not found an effect of familiarity. Collishaw and Hole (2000) found similar sized inversion effects in a recognition task with familiar faces and an old-new memory task with unfamiliar faces. Similarly, Barton et al. (2006) found similar patterns of results for famous and novel faces in eye tracking data during a simple recognition task. They found for both famous and novel faces, the number of fixations and the total scanning duration increased when the image was presented inverted, and there were more fixations to the upper half of faces when upright and more fixations to the lower half of faces when presented inverted for both famous and unfamiliar faces. This mixed evidence suggests that further research is needed to better understand the role of familiarity on the face inversion effect.

As noted throughout the literature covered so far, highly controlled stimuli are used to study the face inversion effect. In an effort to control variables not of interest, the face stimuli often used in experiments are almost always in greyscale, have the background removed, have the majority of the person (the body) cropped out, and often have the face harshly cropped to remove the external features. For example, Hole (1994) showed participants pairs of face stimuli which either had identical or different top halves of the face, paired with different bottom halves of the face. The images were of completely novel faces, were in greyscale and had the body, neck, ears and hairline removed. They found participants were slower at identifying the top halves as being the same when the images were upright compared to inverted. Despite the advancement of time and technology, similar stimuli are used in more recent research. For example, in a study in 2019 by Hadad et al., participants viewed greyscale oval-cropped unfamiliar faces and decided if pairs of simultaneously presented faces were the same or different. They found a larger inversion effect for own-race faces compared to other-race faces in typically developed participants but not autistic participants. These highly controlled stimuli might be specific to the lab and do not resemble how we view faces in day to day life.

Although such stimuli have been used in order to create tightly controlled experiments, this may give rise to a number of issues. For example, in many studies examining the face inversion effect, a control condition is used to compare the effects of inversion on faces with another stimuli class. However, the comparison stimuli do not always undergo equal image manipulations as the face stimuli. For example, Ashworth et al. (2008) contrasted faces with Greebles, artificial non-face objects. The images of faces were greyscale, poor resolution, and cropped into circles removing the entire body and the majority of the external facial features, surrounded by a thick white circle border presented on a black square. In contrast, the images of Greebles were of higher resolution, with no borders or backgrounds, and most notably were not cropped. The authors argued that “using similar cropping for Greebles would have rendered them almost impossible to recognize,” (Ashworth et al., 2008, p. 759). However, this introduces quite a large bias into such experiments, particularly as external features are known to be important for unfamiliar face recognition (Bonner et al., 2003; Ellis et al., 1979). Similar unequal image cropping of face and comparison stimuli can be observed in some of the experiments discussed above. For example, Diamond and Carey (1986) used faces cropped below the chin compared with whole uncropped images of dogs and landscapes. Similarly, Yin (1969) also used images of faces cropped below the chin in comparison to whole uncropped images of houses, aeroplanes and men in motion.

A potential issue with harshly cropping face stimuli is that the removed background might play a role in face recognition. In Mandler’s (1980) classic example, he describes the feeling of familiarity but the difficulty to identify a face when observed in a different context to the one in which the face was originally encountered. Experimental evidence supports Mandler's claim, that face recognition is more accurate in the original context. Gruppuso et al. (2007) showed participants images of faces paired with contexts (e.g. images of buildings, scenery). They then tested participants' memory of the faces which were either shown in the original context, a switched context (one of the previously encountered contexts which was paired with a different face), or a novel context. They found recognition was highest in the original context, decreased for the switched context and lowest for the novel context, lending support to the idea that faces are best recognised in the context they were originally encountered in. This research generalises well to real-life face recognition, as faces are always viewed in front of a background in day-to-day life.

However, in face inversion experiments, the face stimuli often have the background removed from the image. Research by Hayes et al. (2010) demonstrates a potential issue with this. Participants viewed images of people (full colour, cropped just below the shoulders) either with the background removed or present (e.g. a house, a landscape), and were tested for their memory of the faces without backgrounds. Participants were less accurate at recognising the faces which were learnt with a background compared to the faces which were learnt without a background. This demonstrates an issue with face inversion experiments which employ stimuli with the background removed, as this is not how faces are experienced in day-to-day life and the removal of backgrounds has been shown to reduce recognition accuracy.

A further potential issue with the use of tightly controlled stimuli (body cropped out, greyscale, external facial features removed) is that they may not represent how faces are typically viewed in day to day life (people with bodies and uncropped heads against backgrounds), and therefore this might pose a problem to the visual system and the generalisability of the results (Valentine, 1988). For example, as noted throughout, most of the experiments on the FIE crop out the body, however, faces are rarely viewed without a body outside of the lab. Rice et al. (2013) showed in a matching task that when the face does not provide helpful information for identification (e.g. similar-looking images of different people, or dissimilar-looking images of the same person), participants used the body for identification. Performance was unchanged even when the face was removed from the image, supporting the idea that the body can play a role in recognition. Furthermore, one of the controls often applied to face stimuli is to make them greyscale, however, faces are seldom viewed in greyscale outside of the lab. Research shows that colour plays a small but reliable role in recognition. Lee and Perrett (1997) found that recognition of greyscale images of celebrities was slightly but significantly reduced compared to recognition from colour images. Furthermore, Yip and Sinha (2002) found that when information about the shape of a face was degraded, recognition of famous faces was worse from greyscale images compared to full colour images. Moreover, as briefly discussed above, face stimuli in FIE research tend to crop faces into ovals, removing the external features of the head. Again, faces are rarely experienced like this outside of the lab. Devue and de Sena (2023) found that participants were less accurate at recognising celebrities from their internal features compared to head shots which were cropped just above the shoulders leaving the head intact. Controlling stimuli in these ways, removing the colour, body and external features, produces stimuli which do not represent normal day to day visual experience with faces and reduce recognition accuracy.

The potential issues with the tightly controlled stimuli used in face inversion research are examined in this chapter’s experiments. Traditional cropped stimuli which have been use in many previous experiments, where the body and background are removed, are tested directly against uncontrolled images which include the background, body and external facial features. The primary aim of the following 6 experiments was to investigate whether the results of FIE experiments are as a result of overly controlled stimuli and are stimulus-bound, or if they can be found in naturalistic uncontrolled images. They also aim to clarify whether face familiarity has a modulatory effect on the face inversion effect. It was hypothesized that the face inversion effect would be reduced by using whole uncropped images, either because the background would provide semantic cues to identity, improving accuracy, or because the images would be naturalistic and typical of normal visual input and therefore would be processed accurately. The first experiment compared the effect of image cropping (cropped around external features or whole uncropped images) on the face inversion effect. The second experiment investigated the role of semantically meaningful backgrounds on face inversion. The third, fourth and fifth experiments aimed to replicate the findings of experiment 1 and 2 using different tasks. The sixth experiment directly investigated the effect of face familiarity on face inversion.

# Experiment 1: Recognition task with familiar faces and buildings

## Introduction

The aim of this experiment was to use famous faces to test the effect of image cropping on the face inversion effect. This was achieved by comparing traditional highly-controlled stimuli (cropped images: body and background removed) against naturalistic stimuli (whole images: meaningful background and upper body included). In day to day life when encountering familiar faces, the cognitive task we carry out is recognition, therefore a straightforward recognition paradigm was used here, where participants saw images of faces and decided if they recognised them or not. In order to test the face inversion effect – the disproportionate decrease in recognition accuracy for inverted faces *compared* to other stimulus categories – a control condition of buildings was included (Tanaka & Farah, 1993; Yin, 1969). In order to equate the stimuli in the two stimulus classes, images of faces and buildings both included the external features, to avoid the bias introduced in some studies where faces have the external features removed but the control condition stimuli do not (Ashworth et al., 2008). As discussed above, face recognition is more accurate when the face is viewed in its original context (Gruppuso et al., 2007; Hayes et al., 2010). Therefore, images of famous faces in front of semantically-meaningful backgrounds were tested (e.g. Donald Trump in the oval office). It was hypothesized that the face inversion effect would be replicated with the cropped images but reduced with the whole uncropped images as the backgrounds would cue identity and aid recognition.

**Methods**

***Participants***

47 right-handed participants (44 females, 3 males, mean age = 19.13, SD = 0.92) took part in this experiment. 2 additional participants were excluded from the analysis, as one participant got the response keys the wrong way round, and one participant took 56 minutes 38 seconds, more than 3 SDs away from the mean completion time. Participants were University of York Psychology undergraduate students and either received £4 or course credit to compensate their time. All participants had normal or corrected-to-normal vision. Informed consent was gained, and the experiment was approved by the University of York Psychology Ethics Board.

***Stimuli***

32 images of familiar IDs, unfamiliar IDs, familiar buildings and unfamiliar buildings were collected, totalling 128, 1 image per identity / building. A familiarity check was carried out to ensure familiar identities and buildings were known and unfamiliar identities and buildings were unknown (see *Design and Procedure*). UK or international celebrities served as familiar identities, and foreign celebrities served as unfamiliar identities. Images of the familiar and unfamiliar IDs were taken from the internet. Photographs of buildings from York city centre and the University of York served as familiar buildings. These were either buildings close to the Psychology department or well-known to York Psychology students, such as Central Hall or York Minster. Some images were taken from the internet and others were taken by the author L.R.S. Images of buildings from Durham city centre and the University of Durham served as unfamiliar buildings and were taken from the internet. A further 20

images of butterflies were collected for use in an attention check. (See Appendix C for further examples).

All images were cropped to a ratio of 1 : 1.5 and were in full colour. In the whole image condition, the face photographs included the head and upper body, and the background was clearly visible and associated with the person (e.g. Donald Trump in the oval office, Andy Murray on Wimbledon Centre Court). In the cropped image condition, photos were cropped to include only the external features and hair. In the full image condition, the building photographs included the entire building, and the background was clearly visible. In the cropped image condition, photos were cropped around the external contours of the buildings. Each image was edited to create 4 versions: cropped upright, cropped inverted, full image upright, full image inverted, as shown in Figure 3.1.1. The images were fully counterbalanced across participants, so each participant only saw each identity / building once in only 1 of the 4 conditions.

***Design and Procedure***

This experiment had 4 independent variables, each with 2 levels, producing 16 conditions. The stimulus type was either a person or building, the familiarity level was either familiar or unfamiliar, the orientation was either upright or inverted, and the image type was either a whole image or a cropped image. Participants completed 8 trials in each condition, 128 trials in total. The experiment was created using the software PsychoPy (Peirce et al., 2019) and run in-person. Accuracy and reaction times were recorded.

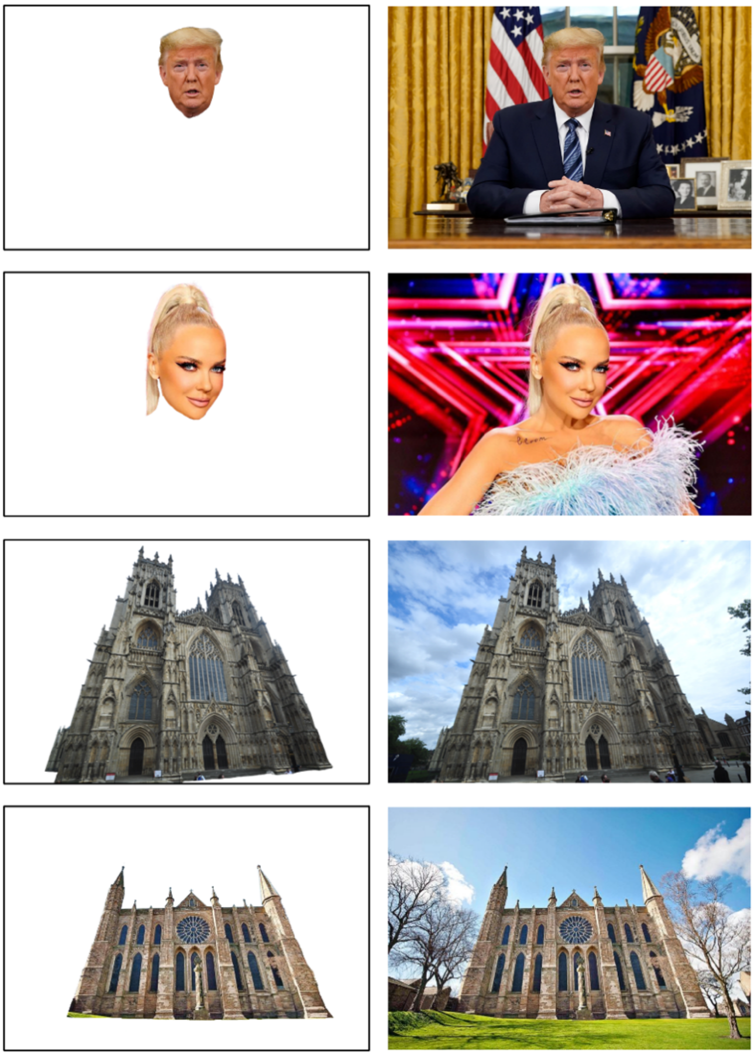


**Figure 3.1.1.** A sample of the stimuli used in Experiment 1. The left column shows cropped stimuli, and the right column shows whole images. The top row represents familiar identities, the second row unfamiliar identities, the third row familiar buildings, and the bottom row unfamiliar buildings.

Participants had to indicate via a button press (1 or 0) if they recognised the person / building or not. At the start of a trial, a fixation cross appeared in the centre of the screen for 0.5s on a grey background. A stimulus then appeared and remained on screen until the participants made a response, as shown in Figure 3.1.2. Images were presented in 1.05 x 0.7 ‘height’ units (the standard size units used in PsychoPy). If a response had not been made after 5s, a prompt briefly appeared telling participants to “please respond quickly.” The order of trials was randomised for each participant and the response keys were counterbalanced across participants.

20 attention check trials were included which were orthogonal to the experimental task to ensure participants paid attention. When an image of a butterfly was presented, participants had to press the “M” key. On average, participants were 95.2% accurate.

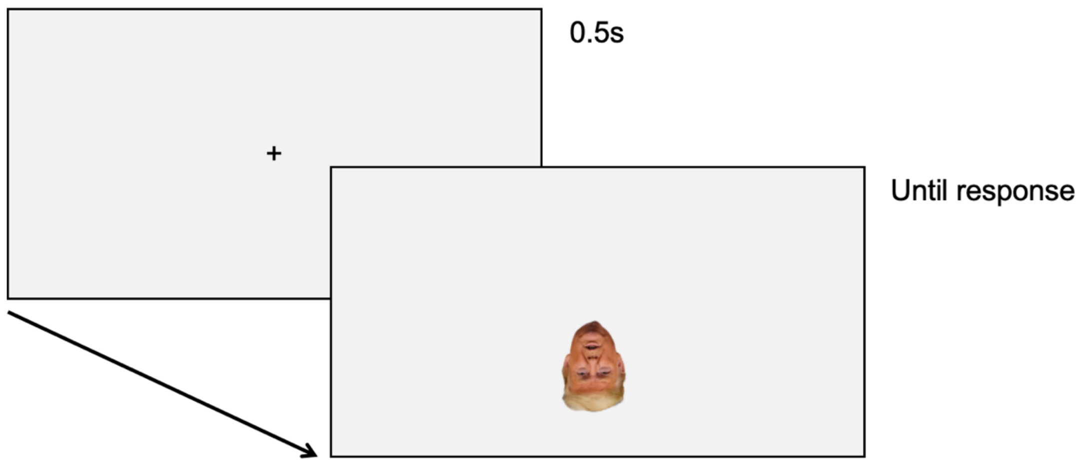



**Figure 3.1.2.** An example trial from Experiment 1. Participants saw a fixation cross followed by a stimulus, and had to indicate via a button press if they recognised the stimulus or not.

After the main experiment, participants completed a familiarity check, to ensure they recognised the familiar identities and buildings, and that they did not recognise the unfamiliar stimuli. Participants were shown each stimulus again, as a whole image in an upright orientation, and were instructed to type in the name of the person /

building. If participants recognised a person / building but could not remember the name, they were instructed to give identifying information, for example a participant could write, “She played Rachel Green on Friends” for Jennifer Anniston, or “Spaceship-style building where graduation takes place” for Central Hall. If participants failed to recognise familiar IDs or buildings, and if they claimed to recognise unfamiliar faces or buildings, those trials were removed prior to analysis; this resulted in 12.8% of trials being eliminated on average for each participant (1 participant had more than 30% of trials removed, 3 participants had more than 20% of trials removed, and 43 participants had less than 20% of trials removed). On average, the experiment lasted approximately 29 and a half minutes.

## Results

*Familiar Identities: Accuracy*

Mean accuracy for familiar face stimuli is shown in Figure 3.1.3. A 2x2 repeated measures ANOVA (whole / cropped face; upright / inverted) was performed on the familiar face accuracy data. Significant main effects of image type ($F(1,46) = 94.51$, $p < .001$, $\eta_p^2 = 0.67$), and orientation ($F(1,46) = 60.54$, $p < .001$, $\eta_p^2 = 0.57$) were found. A significant interaction was found ($F(1,46) = 18.30$, $p < .001$, $\eta_p^2 = 0.29$). Simple main effects tests revealed that the interaction reflected the fact that whole image inversion led to a 6.3% ($p = .001$) decrease in accuracy, whereas cropped face inversion led to a 19.2% ($p < .001$) decrease in accuracy.

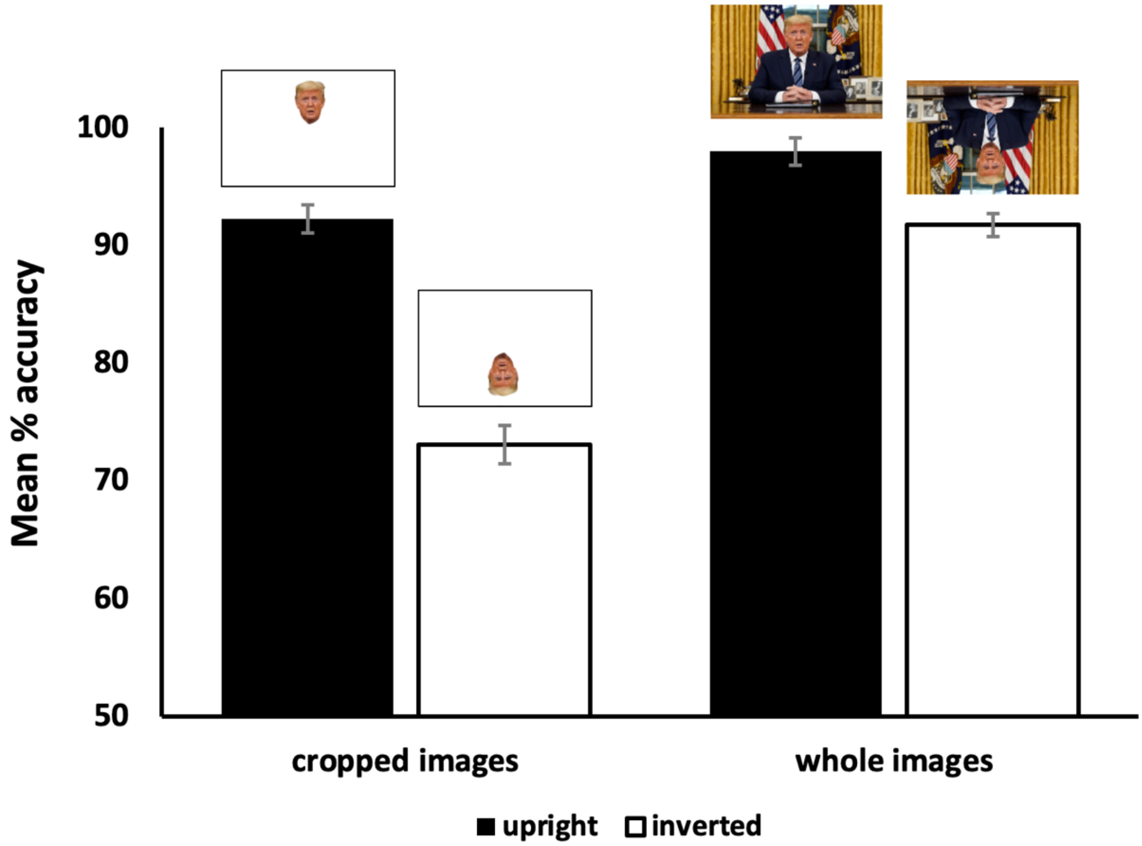



**Figure 3.1.3.** Mean percentage accuracy for recognition of familiar identities shown upright or inverted, and as cropped or whole images. Here and throughout, error bars show within-subject standard error of the mean calculated by the method presented in Cousineau (2005).

*Familiar Identities: Inter-subject variability*

A measure of inter-subject variability was calculated for each condition for the following 6 experiments. If a participant's accuracy was higher in the upright condition compared to the inverted condition, they were classed as showing an inversion effect. If participants' accuracy was lower in the upright condition compared to the inverted condition, they were classed as showing the opposite effect. If their accuracy was equal in the upright and inverted conditions (often reflecting ceiling effects), they were classed as showing no effect. The percentage of participants either showing an effect, showing no effect, or showing the opposite effect was then calculated and is shown in the following pie charts.

For the cropped faces, an inversion effect was observed in 81% of participants (their accuracy was lower in the inverted condition than the upright condition), for 13% no effect was observed (their accuracy was the same in the upright and inverted conditions), and for 6% the opposite effect was observed (their accuracy was higher in the inverted condition compared to upright), as shown in Figure 3.1.4. For the whole images of people, only 36% of participants showed an inversion effect, 57% showed no effect, and 6% showed the opposite effect.

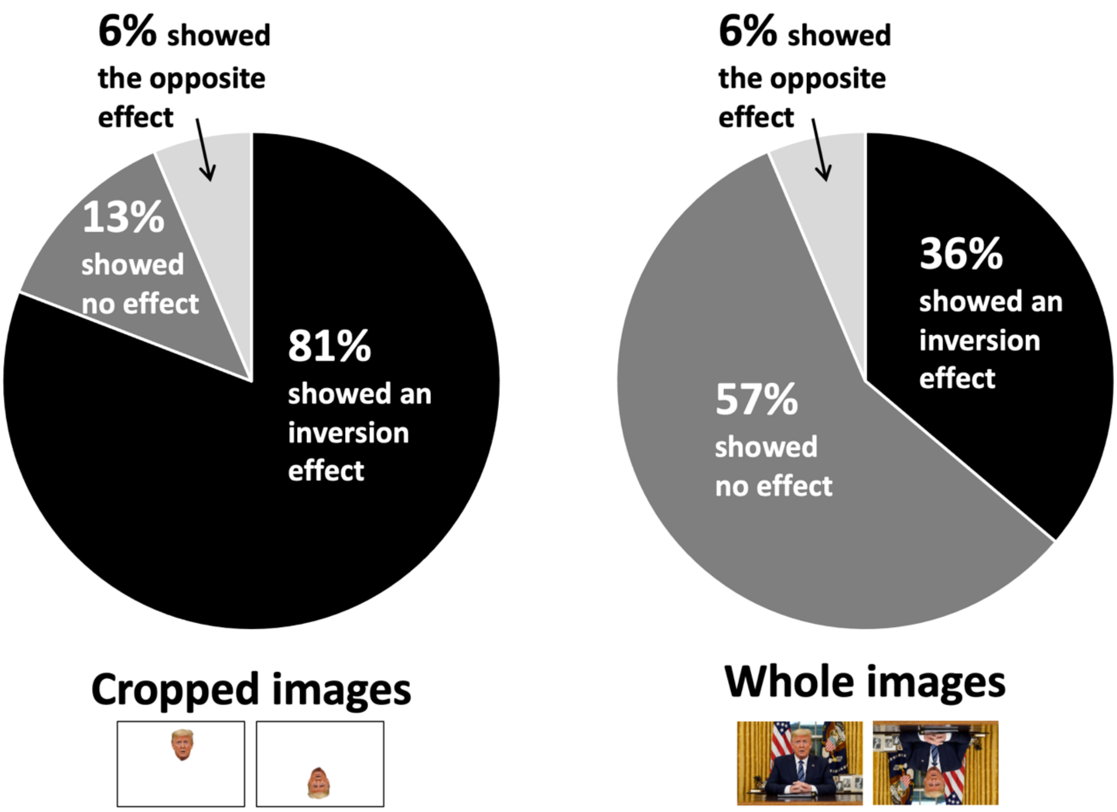


**Figure 3.1.4.** Pie charts displaying the percentage of participants who showed an inversion effect (higher accuracy for upright than inverted), participants who showed no effect (equal accuracy in upright and inverted conditions), and participants who showed the opposite effect (lower accuracy for upright than inverted) for familiar identities shown as cropped or whole images.

*Familiar Identities: RTs*

Figure 3.1.5 shows mean RTs for familiar faces. A 2x2 repeated measures ANOVA (whole / cropped face; upright / inverted) showed no significant main effect of image type ($F(1,46) = 1.79$, $p = .118$, $\eta_p^2 = 0.04$), however, a significant main effect of orientation was found ($F(1,46) = 55.70$, $p < .001$, $\eta_p^2 = 0.55$). No significant interaction was observed ($F(1,46) = 0.30$, $p = .587$, $\eta_p^2 = 0.01$).

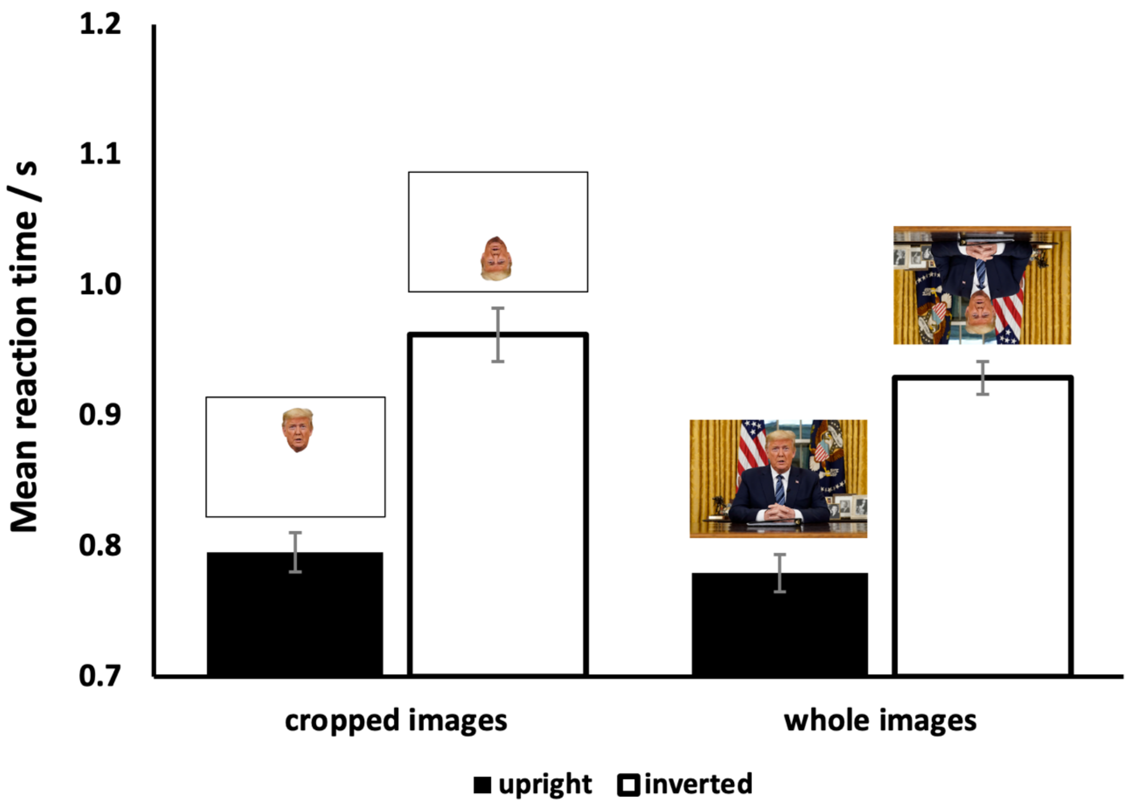


**Figure 3.1.5.** Mean reaction time for recognition of familiar identities shown upright or inverted, and as cropped or whole images.

*Familiar Buildings: Accuracy*

Figure 3.1.6 shows mean accuracy for familiar buildings. A 2x2 repeated measures ANOVA (whole / cropped building; upright / inverted) showed no significant main effect of image type ($F(1,46) = 0.04$, $p = .843$, $\eta_p^2 = 0.00$), however, a significant main effect of orientation ($F(1,46) = 7.44$, $p = .009$, $\eta_p^2 = 0.14$) was found. No significant interaction was observed ($F(1,46) = 0.04$, $p = .838$, $\eta_p^2 = 0.00$).

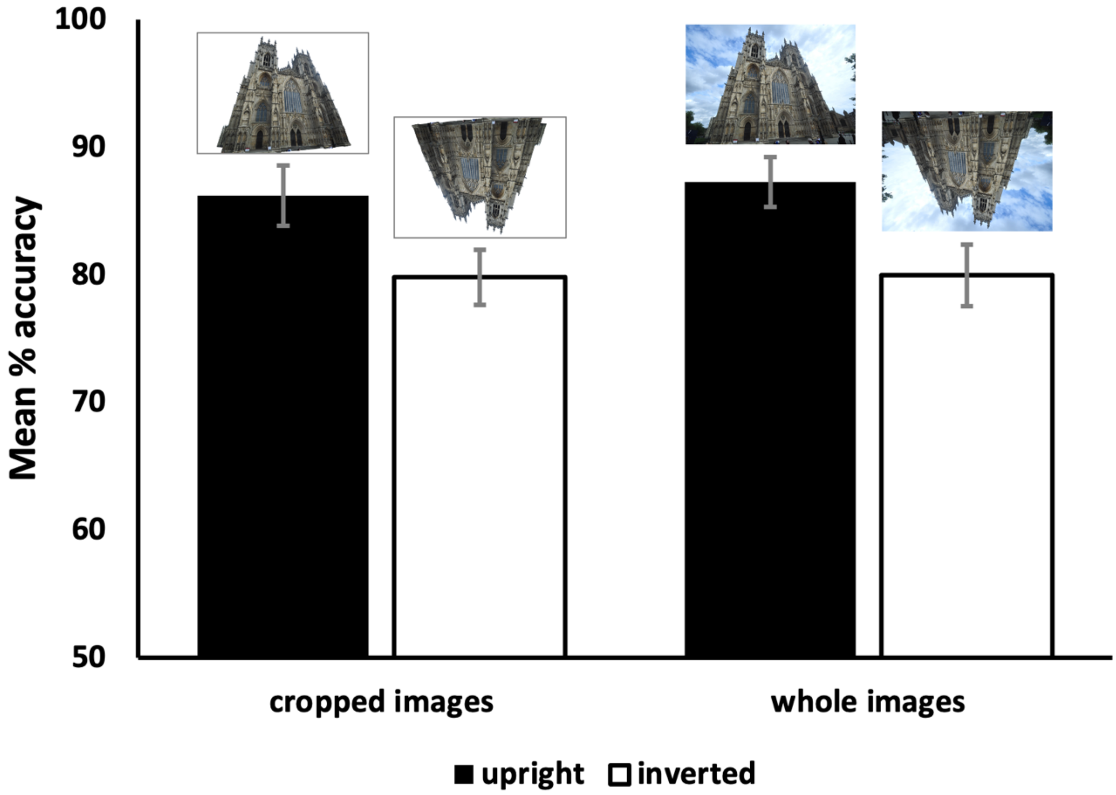



**Figure 3.1.6.** Mean percentage accuracy for recognition of familiar buildings shown upright or inverted, and as cropped or whole images.

*Familiar Buildings: Inter-subject variability*

For the cropped images, 53% of participants showed an inversion effect, 21% showed no effect and 26% showed the opposite effect, as shown in Figure 3.1.7. Similarly, for whole images, 55% showed an inversion effect, 26% showed no effect, and 19% showed the opposite effect.

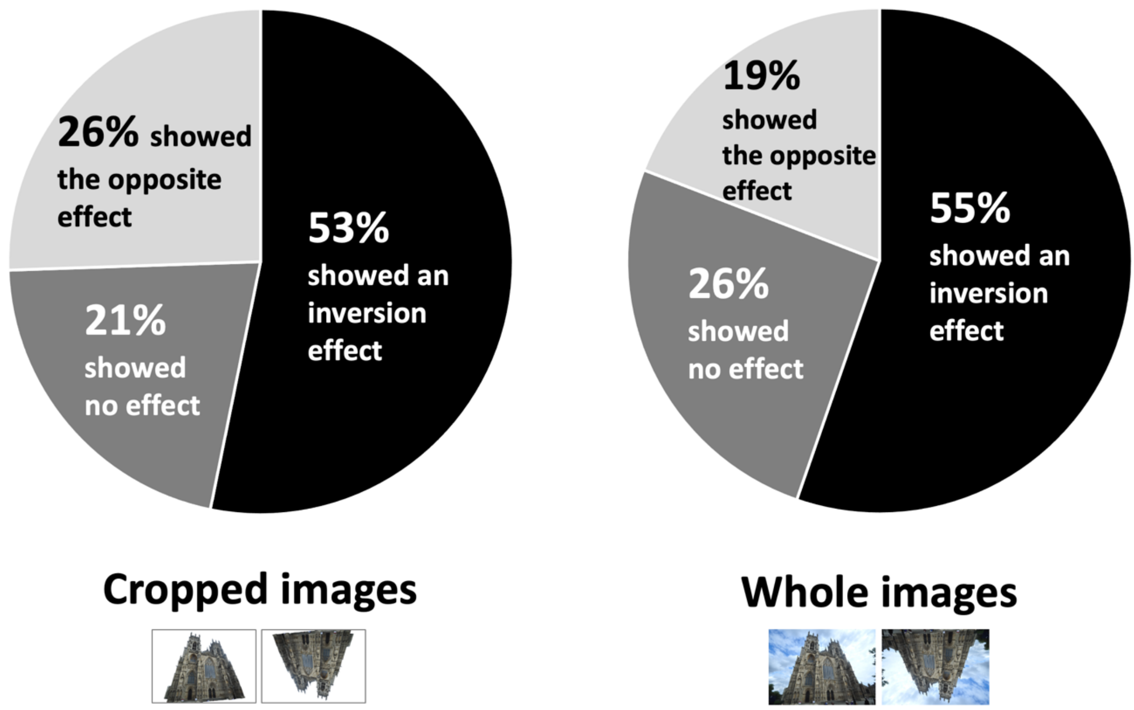



**Figure 3.1.7.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for familiar buildings shown as cropped or whole images.

*Familiar Buildings: RTs*

Figure 3.1.8 shows mean RTs for familiar buildings. A 2x2 repeated measures ANOVA (whole / cropped building; upright / inverted) showed no significant main effect of image type (F(1,46) = 0.95, $p$ = .335, $\eta_p^2$ = 0.02), however, a significant main effect of orientation was found (F(1,46) = 8.48, $p$ = .006, $\eta_p^2$ = 0.16). No significant interaction was observed (F(1,46) = 0.17, $p$ = .669, $\eta_p^2$ = 0.00).

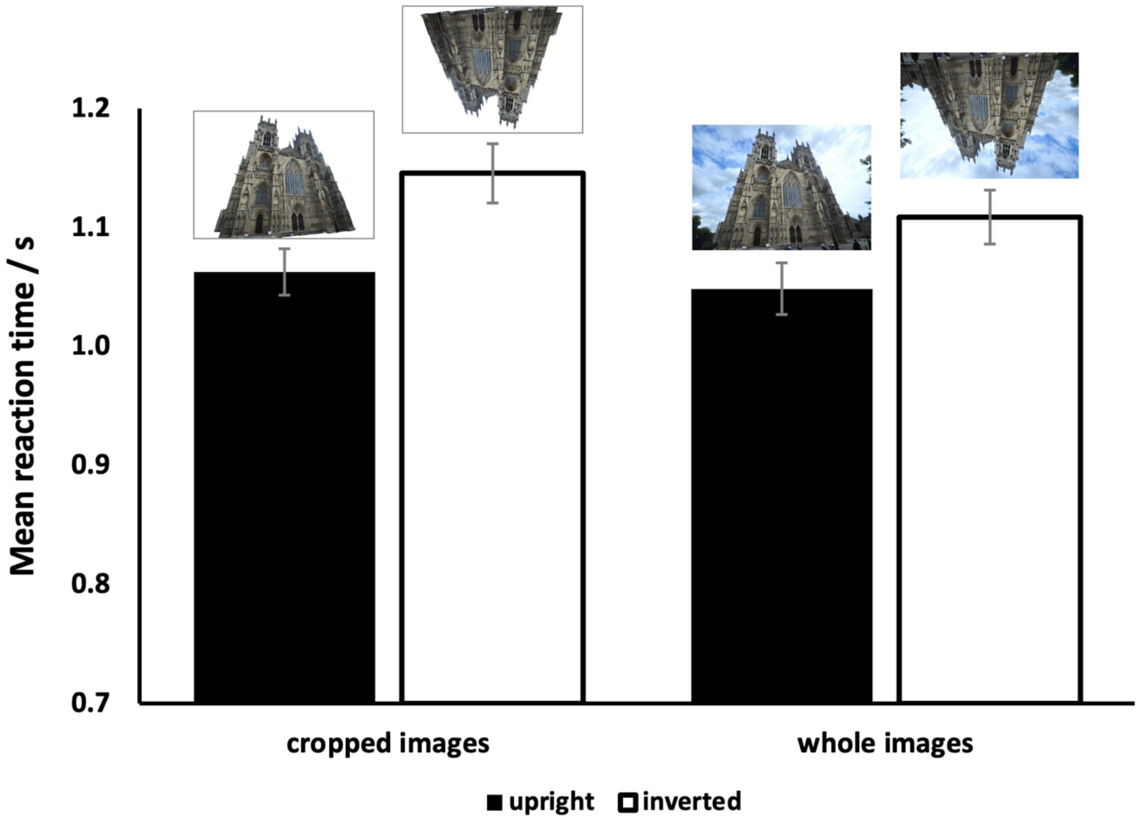



**Figure 3.1.8.** Mean reaction time for recognition of familiar buildings shown upright or inverted, and as cropped or whole images.

*Unfamiliar Identities: Accuracy*

Mean accuracy for familiar face stimuli is shown in Figure 3.1.9. A 2x2 repeated measures ANOVA (whole / cropped face; upright / inverted) showed no significant main effect of image type (F(1,46) = 2.50, $p$ = .121, $\eta_p^2$ = 0.05), however, a significant main effect of orientation (F(1,46) = 22.77, $p$ < .001, $\eta_p^2$ = 0.33) was found. No significant interaction was observed (F(1,46) = 0.62, $p$ = .435, $\eta_p^2$ = 0.01).

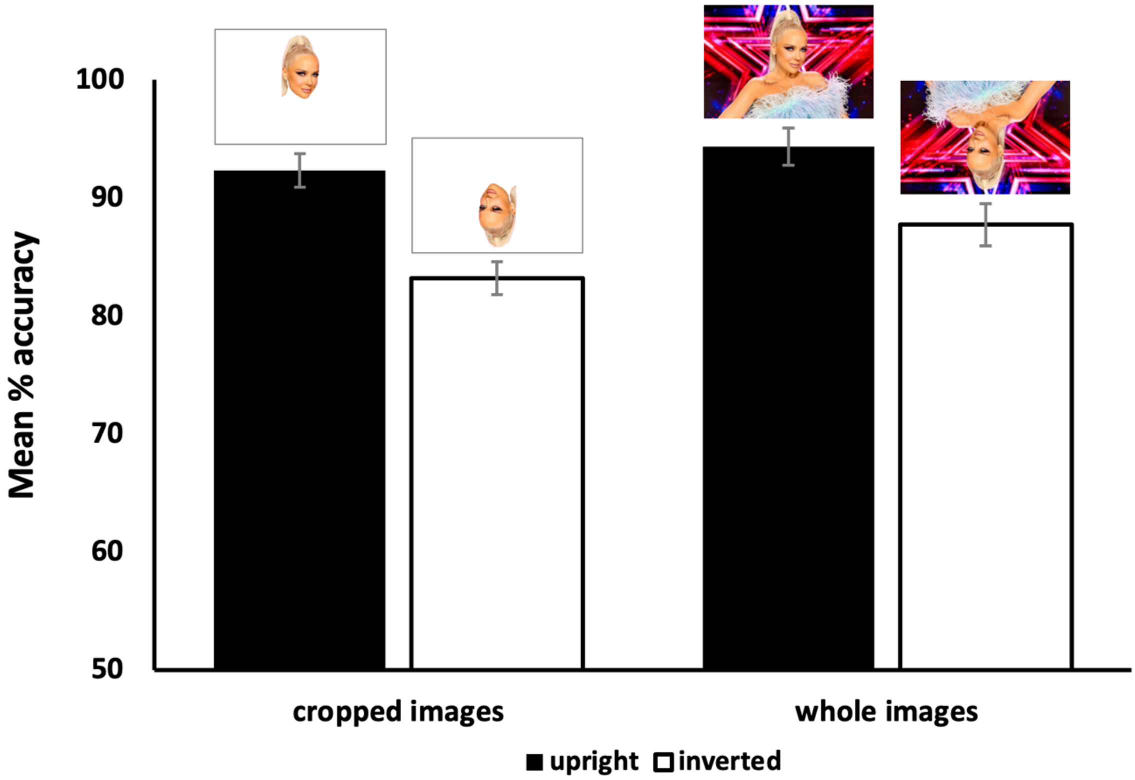



**Figure 3.1.9.** Mean percentage accuracy for recognition of unfamiliar identities shown upright or inverted, and as cropped or whole images.

*Unfamiliar Identities: Inter-subject variability*

For the cropped images, 49% of participants showed an inversion effect, 36% showed no effect and 15% showed the opposite effect, as shown in Figure 3.1.10. Similarly, for whole images, 45% showed an inversion effect, 45% showed no effect, and 11% showed the opposite effect.

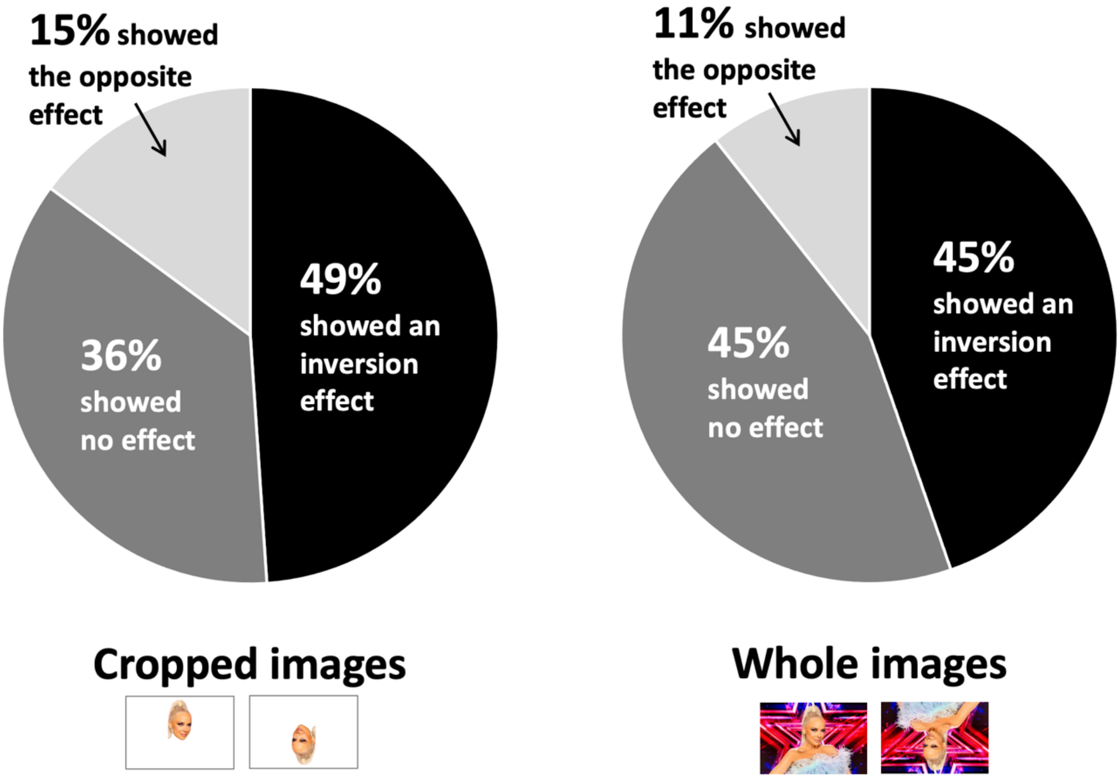



**Figure 3.1.10.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for unfamiliar identities shown as cropped or whole images.

*Unfamiliar Identities: RTs*

Figure 3.1.11 shows mean RTs for unfamiliar identities. A 2x2 repeated measures ANOVA (whole / cropped face; upright / inverted) showed no significant main effect of image type ($F(1,46) = 1.56$, $p = .218$, $\eta_p^2 = 0.03$), however, a significant main effect of orientation was found ($F(1,46) = 24.96$, $p < .001$, $\eta_p^2 = 0.35$). No significant interaction was observed ($F(1,46) = 0.41$, $p = .526$, $\eta_p^2 = 0.01$).

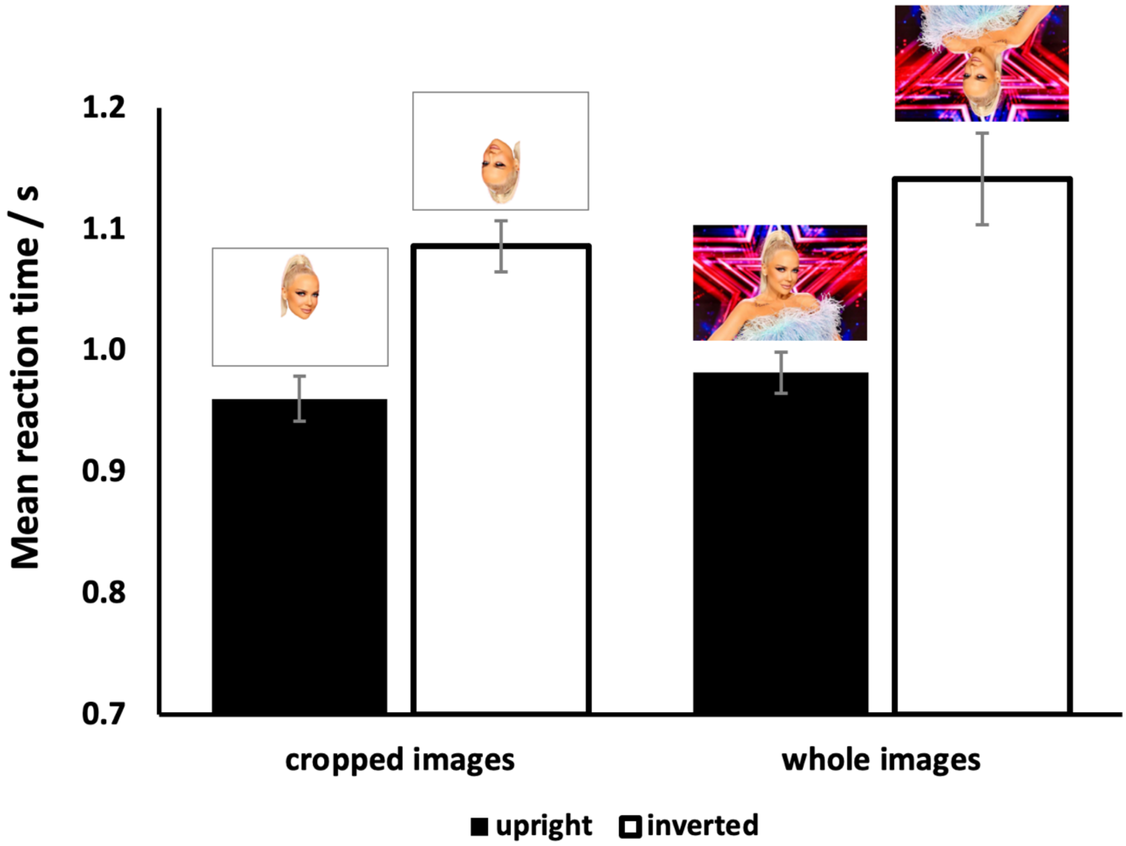



**Figure 3.1.11.** Mean reaction time for recognition of unfamiliar identities shown upright or inverted, and as cropped or whole images.

*Unfamiliar Buildings: Accuracy*

Figure 3.1.12 shows mean accuracy for familiar buildings. A 2x2 repeated measures ANOVA (whole / cropped building; upright / inverted) showed no significant main effects of image type ($F(1,46) = 0.14$, $p = .715$, $\eta_p^2 = 0.00$), or orientation ($F(1,46) = 1.28$, $p = .264$, $\eta_p^2 = 0.03$). No significant interaction was observed ($F(1,46) = 0.01$, $p = .908$, $\eta_p^2 = 0.00$).


**Figure 3.1.12.** Mean percentage accuracy for recognition of unfamiliar buildings shown upright or inverted, and as cropped or whole images.

*Unfamiliar Buildings: Inter-subject variability*

For the cropped images, 36% of participants showed an inversion effect, 34% showed no effect and 30% showed the opposite effect, as shown in Figure 3.1.13. Similarly, for whole images, 36% showed an inversion effect, 32% showed no effect, and 32% showed the opposite effect.

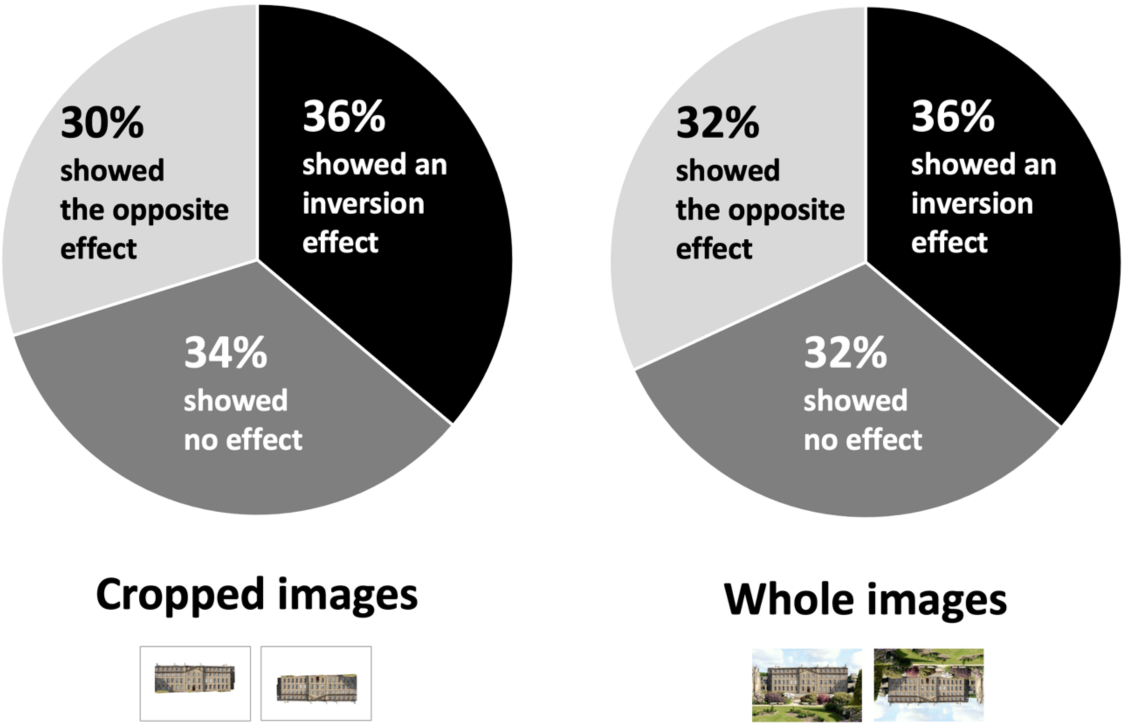



**Figure 3.1.13.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for unfamiliar buildings shown as cropped or whole images.

*Unfamiliar Buildings: RTs*

Figure 3.1.14 shows mean RTs for unfamiliar buildings. A 2x2 repeated measures ANOVA (whole / cropped building; upright / inverted) showed no significant main effects of image type ($F(1,46) = 0.69$, $p = .410$, $\eta_p^2 = 0.02$), or orientation ($F(1,46) = 2.45$, $p = .125$, $\eta_p^2 = 0.05$) were found. No significant interaction was observed ($F(1,46) = 1.44$, $p = .237$, $\eta_p^2 = 0.03$).

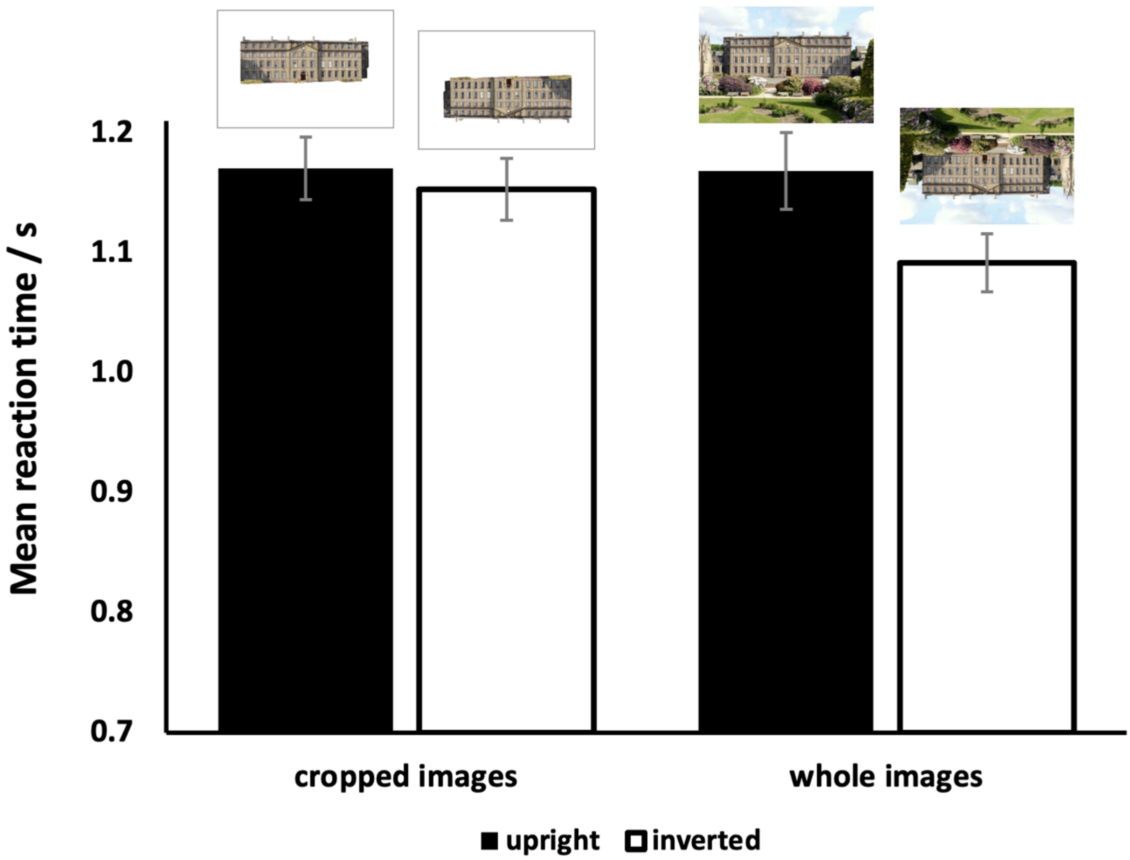



**Figure 3.1.14.** Mean reaction time for recognition of unfamiliar buildings shown upright or inverted, and as cropped or whole images.

**Further Analysis**

In order to formally compare the face inversion effect, the larger effect of inversion on faces compared to buildings, 3-way ANOVAs were run.

*Familiar faces and buildings: Accuracy*

A 2x2x2 repeated measures ANOVA (stimulus type: faces / buildings; image type: whole / cropped; orientation: upright / inverted) was performed on the accuracy data for familiar faces and familiar buildings. Significant main effects of stimulus type

($F(1,46) = 8.32$, $p = .006$, $\eta_p^2 = 0.15$), image type ($F(1,46) = 16.65$, $p < .001$, $\eta_p^2 = 0.27$), and orientation ($F(1,46) = 46.19$, $p < .001$, $\eta_p^2 = 0.50$) were found. The stimulus type * image type two-way interaction ($F(1,46) = 12.07$, $p = .001$, $\eta_p^2 = 0.21$), and image type * orientation ($F(1,46) = 7.43$ $p = .009$, $\eta_p^2 = 0.14$) were significant, however, the stimulus type * orientation interaction was not significant ($F(1,46) = 3.54$, $p = .066$, $\eta_p^2 = 0.07$). The three-way interaction between stimulus type, image type and orientation was significant ($F(1,46) = 5.29$, $p = .026$, $\eta_p^2 = 0.10$).

*Familiar faces and buildings: RTs*

A 2x2x2 repeated measures ANOVA (stimulus type: faces / buildings; image type: whole / cropped; orientation: upright / inverted) was performed on the RT data for familiar faces and familiar buildings. Significant main effects of stimulus type ($F(1,46) = 73.47$, $p < .001$, $\eta_p^2 = 0.62$) and orientation ($F(1,46) = 51.15$ $p < .001$, $\eta_p^2 = 0.53$) were found, however there was no significant main effect of image type ($F(1,46) = 1.99$, $p = .165$, $\eta_p^2 = 0.04$). The stimulus type * orientation two-way interaction was significant ($F(1,46) = 6.99$, $p = .011$, $\eta_p^2 = 0.13$), however the stimulus type * image type ($F(1,46) = 0.00$, $p = .973$, $\eta_p^2 = 0.00$) and the image type * orientation interaction was not significant ($F(1,46) = 0.42$ $p = .518$, $\eta_p^2 = 0.01$). The three-way interaction between stimulus type, image type and orientation was not significant ($F(1,46) = 0.01$, $p = .924$, $\eta_p^2 = 0.00$).

*Unfamiliar faces and buildings: Accuracy*

A 2x2x2 repeated measures ANOVA (stimulus type: faces / buildings; image type: whole / cropped; orientation: upright / inverted) was performed on the accuracy data for unfamiliar faces and unfamiliar buildings. A significant main effect of orientation was found ($F(1,46) = 15.44$ $p < .001$, $\eta_p^2 = 0.25$), however the main effects of stimulus type ($F(1,46) = 0.64$, $p = .430$, $\eta_p^2 = 0.01$) and image type ($F(1,46) = 1.30$, $p = .260$, $\eta_p^2 = 0.03$) were not significant. The stimulus type * orientation two-way interaction was significant ($F(1,46) = 6.90$, $p = .012$, $\eta_p^2 = 0.13$), however the stimulus type * image type ($F(1,46) = 1.90$, $p = .174$, $\eta_p^2 = 0.04$) and the image type * orientation interaction was not significant ($F(1,46) = 0.18$ $p = .670$, $\eta_p^2 = 0.00$). The three-way interaction between stimulus type, image type and orientation was not significant ($F(1,46) = 0.33$, $p = .567$, $\eta_p^2 = 0.01$).

*Unfamiliar faces and buildings: RTs*

A 2x2x2 repeated measures ANOVA (stimulus type: faces / buildings; image type: whole / cropped; orientation: upright / inverted) was performed on the RT data for unfamiliar faces and unfamiliar buildings. Significant main effects of stimulus type ($F(1,46) = 14.29$, $p < .001$, $\eta_p^2 = 0.24$) and orientation ($F(1,46) = 4.15$ $p = .047$, $\eta_p^2 = 0.08$) were found, however there was no significant main effect of image type ($F(1,46) = 0.04$, $p = .837$, $\eta_p^2 = 0.00$). The stimulus type * orientation two-way interaction was significant ($F(1,46) = 29.28$, $p < .001$, $\eta_p^2 = 0.39$), however the stimulus type * image type ($F(1,46) = 1.33$, $p = .254$, $\eta_p^2 = 0.03$) and the image type * orientation interaction was not significant ($F(1,46) = 0.11$ $p = .739$, $\eta_p^2 = 0.00$). The three-way interaction between stimulus type, image type and orientation was not significant ($F(1,46) = 1.74$, $p = .194$, $\eta_p^2 = 0.04$).

**Discussion**

For familiar cropped faces, I find the classic face inversion effect: faces suffer much larger inversion effects compared to buildings (faces 19.2%, buildings 6.4%). However, with the whole images, faces and buildings suffer similar inversion effects (faces 6.3%, buildings 7.3%), in accordance with the hypothesis. The condition most severely affected by inversion was the cropped famous faces. The same pattern can be seen in the inter-subject variability results (the percentage of participants that do / do not show inversion effects shown in the pie charts). With famous faces, for cropped images 81% of participants showed an inversion effect compared to 36% for whole images. However, for unfamiliar faces, familiar buildings, and unfamiliar buildings, the cropped images and whole images produced near-identical percentages of participants showing inversion effects.

Some of the results in this experiment may have suffered from ceiling effects. For example, upright cropped famous faces were recognised with 92.2% accuracy and upright whole images at 98.0%. Therefore, looking also at the reaction time data can help to interpret the results. Reaction times were fastest for famous faces overall, which is consistent with the very high accuracy scores. For famous faces, the RT data did not show an interaction, with both cropped and whole images showing similar increases in RTs when inverted. One possible explanation could be that if the accuracy scores were brought down from ceiling, e.g. if the data were transformed, then they might reveal similar sized decreases in accuracy when the image is inverted for both cropped and whole images, consistent with the pattern found in the RT data. Another possibility is that image cropping may affect accuracy scores more than reaction times, and therefore any conclusions drawn would need to reflect that the effect of image cropping only affects one measure of the FIE.

The difference in accuracy between cropped and whole images of inverted familiar identities is striking. Inverted cropped images of familiar faces suffered a further decrease of 12.9 percentage points in comparison to inverted whole images of familiar IDs. This experiment clearly shows that using whole naturalistic images of familiar faces produces considerably different results compared to the cropped faces often used in FIE experiments. In fact, inverting whole images of familiar faces produced similar decreases in accuracy as inverting images of buildings. The face inversion effect, the disproportionately larger decrease in recognition accuracy of faces compared to non-face objects, is not found in this experiment using whole images. This represents an unexpected finding, as the face inversion effect is a widely reproduced result (McKone & Yovel, 2009), and it is taken as evidence that faces are special (Yin, 1969).

What then might be causing the large differences in inversion between the whole and cropped images of familiar faces? One explanation is the presence or absence of backgrounds. The images were chosen to have the familiar identities in front of meaningful backgrounds linked to the person, for example, Queen Elizabeth II sat on a golden throne in an ornate golden room. Previous research by Gruppuso et al. (2007) showed that identities are most accurately recognised when they are presented in their original context, therefore the semantically meaningful backgrounds in the whole image condition may have cued identity, aiding recognition and reducing the effect of inversion. Furthermore, Hayes et al. (2010) found that recognition of identities is hindered when the background of an image is removed. Therefore, recognition of the identities in the cropped condition may have been inhibited by both inversion and the cropping removing the background. The presence of semantically rich backgrounds in the whole images and the removal of the background in the cropped images may have caused the large differences in recognition accuracy of inverted cropped and whole images.

However, from this experiment alone, the exact factor which is restored in whole images cannot be pinpointed. In the following experiment, this is addressed directly by including an intermediate condition where the background is removed, but the external features of the whole body remain intact.

# Experiment 2: Recognition task with familiar cropped faces, whole bodies and whole images

### Introduction

The previous experiment found larger inversion effects for cropped faces (19.2%) compared to whole images (6.3%). In order to test if the removal of the background caused the larger decrease in accuracy, an intermediate condition is added in this experiment where the background is removed, but the full external contour of the person is kept intact. The same procedure is used as in Experiment 1, however, cropped faces, whole bodies and whole images are directly compared.

One possible explanation of the results found in Experiment 1 is that in the whole image condition, the background cued identity (Gruppuso et al., 2007) reducing the effect of inversion, and the removal of the background in the cropped faces condition inhibited recognition (Hayes et al., 2010). However, an alternative explanation which cannot be ruled out based on Experiment 1 alone, is that the large inversion effects were due to large distortions of the test stimuli away from what they represent in real life. The cropped faces had the background and entire body cropped out of the image and were inverted, however, this is not how faces are experienced in real life and may have led to the large reduction in accuracy. This experiment aims to test between these 2 hypotheses by adding in a *whole body* condition. If the reduced inversion effect found in Experiment 1 with whole images was due to the semantically meaningful backgrounds cueing identity, then it is hypothesized that cropped faces and whole bodies will suffer similar size inversion effects, which would be much larger compared to inversion effects found using whole images. However, if the large inversion effects found when inverting cropped faces in Experiment 1 were due to the abnormal cropping, then it is hypothesized that cropped faces will produce the largest inversion effect, followed by whole bodies, and whole images will produce the smallest inversion effects.

## Methods

### *Participants*

67 right-handed participants (40 females, 25 males, 2 participants identified as non-binary, mean age = 24.65, SD = 3.36, 1 participant gave an inaccurate DOB and therefore their data is not included in the mean age or SD) took part in this experiment. 5 additional participants were excluded from the analysis (1 participant produced 2 complete data sets, 2 participants experienced data recording issues, 1 participant got the response keys the wrong way round, and 1 participant had missing cases in their RT data). Participants were recruited from Prolific and were paid £5 to compensate their time. All participants self-reported as being right-handed, currently residing in the United Kingdom, had lived in the UK for at least 10 years, and had normal or corrected-to-normal vision. Informed consent was gained, and the experiment was approved by the University of York Psychology Ethics Board.

### *Stimuli*

The majority of the 32 familiar and unfamiliar IDs from Experiment 1 were used. For 11 of the famous identities, new photographs were chosen which would be more suitable for the whole body cropping condition, e.g. a photograph of Leonardo DiCaprio holding an Oscar in front of him would result in a whole body cropped photo with an Oscar-shaped hole in his body; therefore this was substituted for an image of him standing in front of a large Oscar statue at the Oscars for example. For 2 of the unfamiliar identities, new images were chosen, and a further 7 unfamiliar identities were replaced with new unfamiliar identities for the same reasons. A further 16 famous and 16 unfamiliar new identities were found, producing a total of 48 familiar and 48 unfamiliar identities. All images were cropped in the same way as in Experiment 1, including the whole image or cropping around the head, but an additional condition was created whereby the images were cropped around the entire body including the head, as shown in Figure 3.2.1. Each image was edited to create 6 versions: cropped head upright, cropped head inverted, whole body upright, whole body inverted, whole image upright and whole image inverted. The images were fully counterbalanced across participants, so each participant only saw each identity once in only 1 of the 6 conditions. The same 20 images of butterflies from Experiment 1 were used again in an attention check. (See Appendix D for further examples).

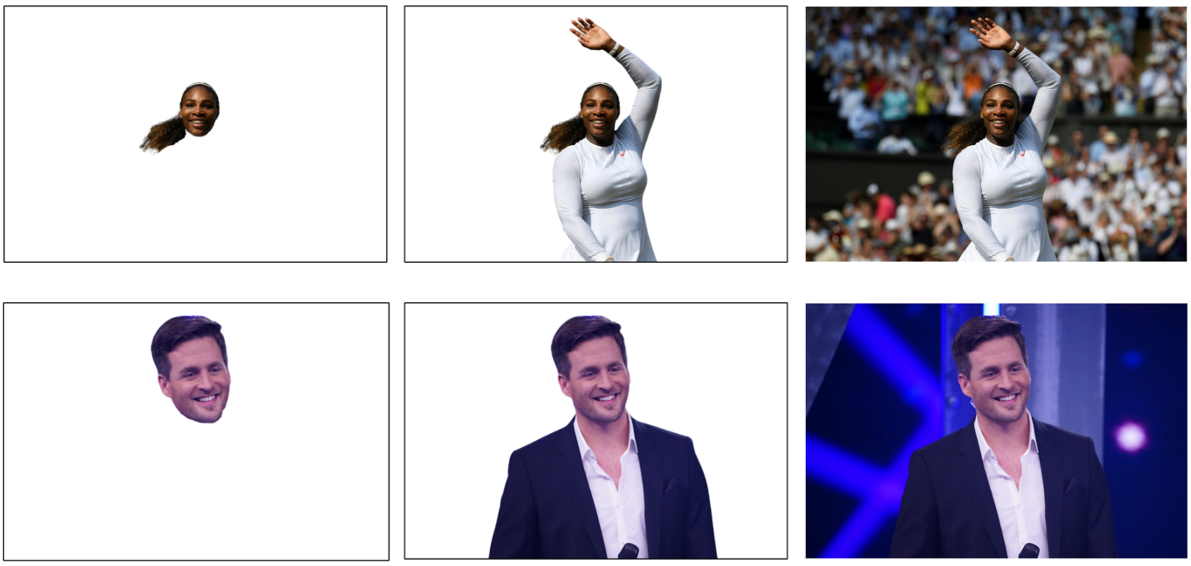


**Figure 3.2.1.** A sample of the stimuli used in Experiment 2. The top row shows familiar identities, and the bottom row shows unfamiliar identities. The left column

demonstrates cropped heads, the middle column whole bodies and the right column whole images.

***Design and Procedure***

This experiment had 3 independent variables. The image type was either a whole image, cropped around the whole body or cropped around the head, the familiarity level was either familiar or unfamiliar, and the orientation was either upright or inverted. Participants completed 8 trials in each condition, 96 trials in total. Accuracy and reaction time were recorded.

The experimental procedure was identical to Experiment 1. Participants completed a recognition task (see Figure 3.2.2) indicating if they recognised the identities or not. Again, intermixed were 20 attention check trials where participants pressed “M” when presented with an image of a butterfly. On average, participants were 96.3% accurate. A familiarity check again followed the main experiment to remove any unrecognized familiar IDs and supposedly recognised unfamiliar faces whereby participants saw the whole upright images again and had to type in their name or identifying information. This resulted in 9.6% of trials being eliminated on average for each participant (1 participant had more than 30% of trials removed, 4 participants had more than 20% of trials removed, 62 participants had less than 20% of trials removed). On average, the experiment lasted approximately 28 minutes.

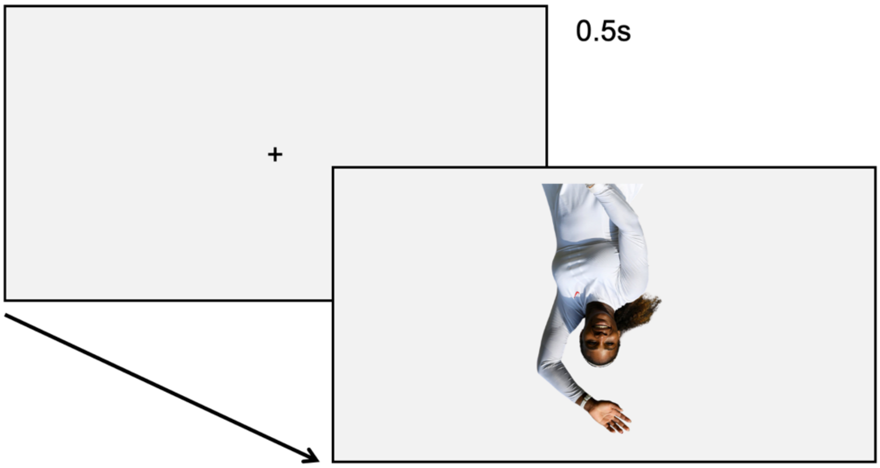



**Figure 3.2.2.** An example trial from Experiment 2. Participants saw a fixation cross followed by a stimulus, and had to indicate via a button press whether they recognised the identity or not.

**Results**

*Familiar Identities: Accuracy*

Mean accuracy for familiar face stimuli is shown in Figure 3.2.3. A 2x3 repeated measures ANOVA (whole image / whole body / cropped head; upright / inverted) was performed on the familiar identity accuracy data. The image type ($p$ = .060) and interaction ($p$ = .066) data met the assumption of sphericity. Significant main effects of image type (F(2,132) = 31.44, $p$ < .001, $\eta_p^2$ = 0.32), and orientation (F(1,66) = 174.73, $p$ < .001, $\eta_p^2$ = 0.73) were found. Bonferroni corrected pairwise comparisons revealed that the main effect of image type was driven by the cropped heads being recognised significantly less accurately than the whole images ($p$ < .001) and whole bodies ($p$ < .001). There was no significant difference between whole images and whole bodies ($p$ = .453). A significant interaction was also found (F(2,132) = 6.57, $p$ =

.002, $\eta_p^2 = 0.09$). Simple main effects tests revealed that the interaction reflected the fact that whole image and whole body inversion led to a 15.1% ($p < .001$) and 16.3% ($p < .001$) decrease in accuracy respectively, whereas cropped face inversion led to a 27.6% ($p < .001$) decrease in accuracy.


**Figure 3.2.3.** Mean percentage accuracy for recognition of familiar identities shown upright or inverted, and as cropped heads, whole bodies or whole images.

*Familiar Identities: Inter-subject variability*

For the cropped faces, an inversion effect was observed in 76% of participants (i.e. their accuracy was lower in the inverted condition than the upright condition), for 15% no effect was observed (their accuracy was the same in the upright and inverted conditions), and for 9% the opposite effect was observed (their accuracy was higher in the inverted condition compared to upright), as shown in Figure 3.2.4. For the whole bodies, 61% of participants showed an inversion effect, 30% showed no effect, and 9% showed the opposite effect. Similarly, for the whole images, 67% of participants showed an inversion effect, 30% showed no effect, and 3% showed the opposite effect.

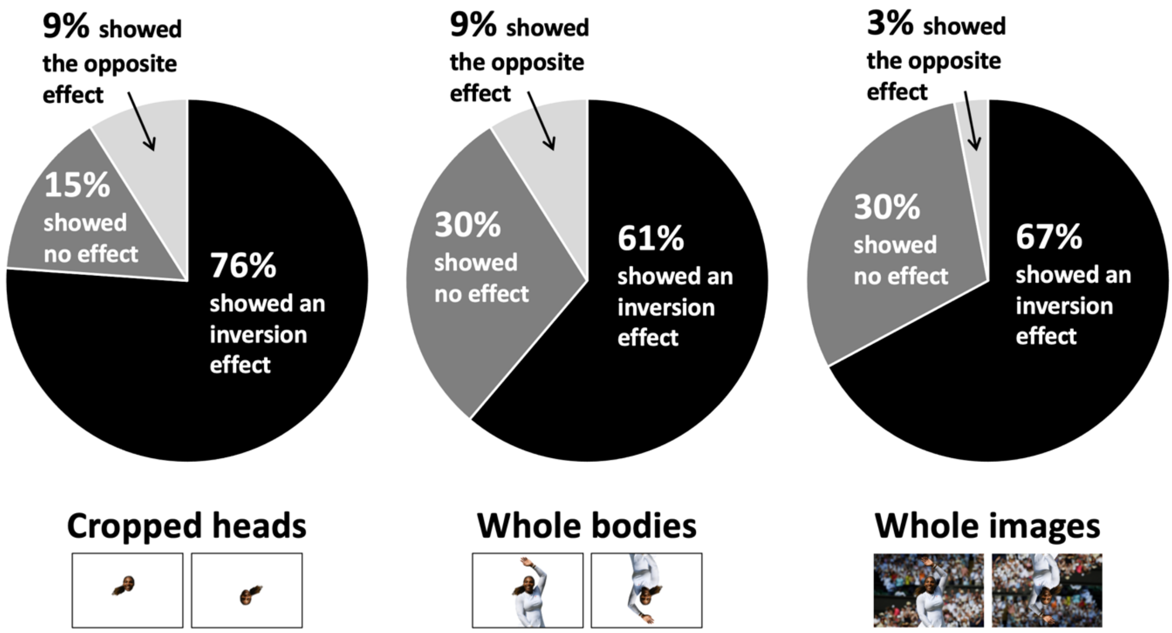



**Figure 3.2.4.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for familiar identities shown as cropped heads, whole bodies or whole images.

*Familiar Identities: RTs*

Figure 3.2.5 shows mean RTs for familiar faces. A 2x3 repeated measures ANOVA (whole image / whole body / cropped head; upright / inverted) was performed on the familiar identity RT data. The image type data did not violate the assumption of sphericity ($p$ = .211). However, the interaction did violate the assumption of sphericity ($p$ < .001), therefore a Greenhouse Geisser correction was applied to the interaction. Significant main effects of image type (F(2,132) = 10.44, $p$ < .001, $\eta_p^2$ = 0.14), and orientation (F(1,66) = 92.10, $p$ < .001, $\eta_p^2$ = 0.58) were found. Bonferroni corrected pairwise comparisons revealed that the main effect of image type was driven by the cropped heads being recognised significantly slower than the whole images ($p$ = .004) and whole bodies ($p$ < .001). There was no significant difference between whole images and whole bodies ($p$ > .999). A significant interaction was also found (F(1.63, 107.46) = 5.21, $p$ = .011, $\eta_p^2$ = 0.07). Simple main effects tests revealed that the interaction reflected the fact that whole image and cropped body inversion both led to a 0.16s increase in RTs (both $p$ < .001), whereas cropped face inversion led to a 0.29s ($p$ < .001) increase in RTs.

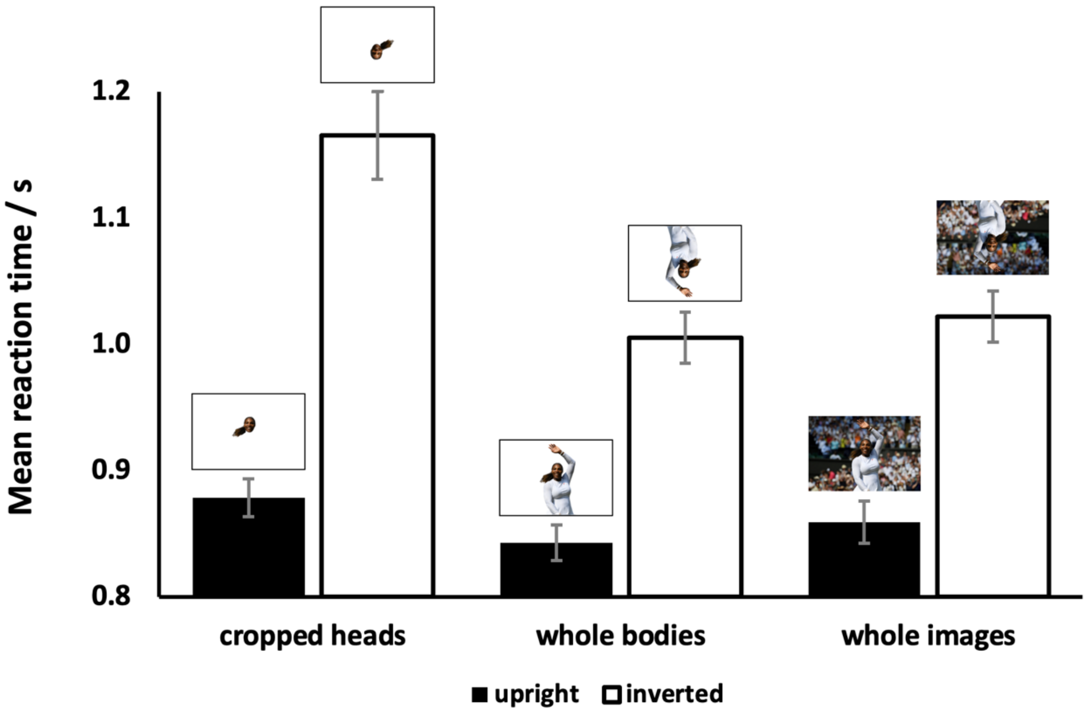



**Figure 3.2.5.** Mean reaction times for recognition of familiar identities shown upright or inverted, and as cropped heads, whole bodies or whole images.

*Unfamiliar Identities: Accuracy*

Figure 3.2.6 shows mean accuracy for unfamiliar identities. A 2x3 repeated measures ANOVA (whole image / whole body / cropped head; upright / inverted) was performed on the unfamiliar face accuracy data. The image type data ($p$ = .627) and interaction ($p$ = .063) met the assumption of sphericity. A significant main effect of orientation was found ($F(1,66) = 19.02$, $p < .001$, $\eta_p^2 = 0.22$), however, there was no significant main effect of image type ($F(2,132) = 1.22$, $p = .297$, $\eta_p^2 = 0.02$). No significant interaction was observed ($F(2,132) = 0.60$, $p = .553$, $\eta_p^2 = 0.01$).

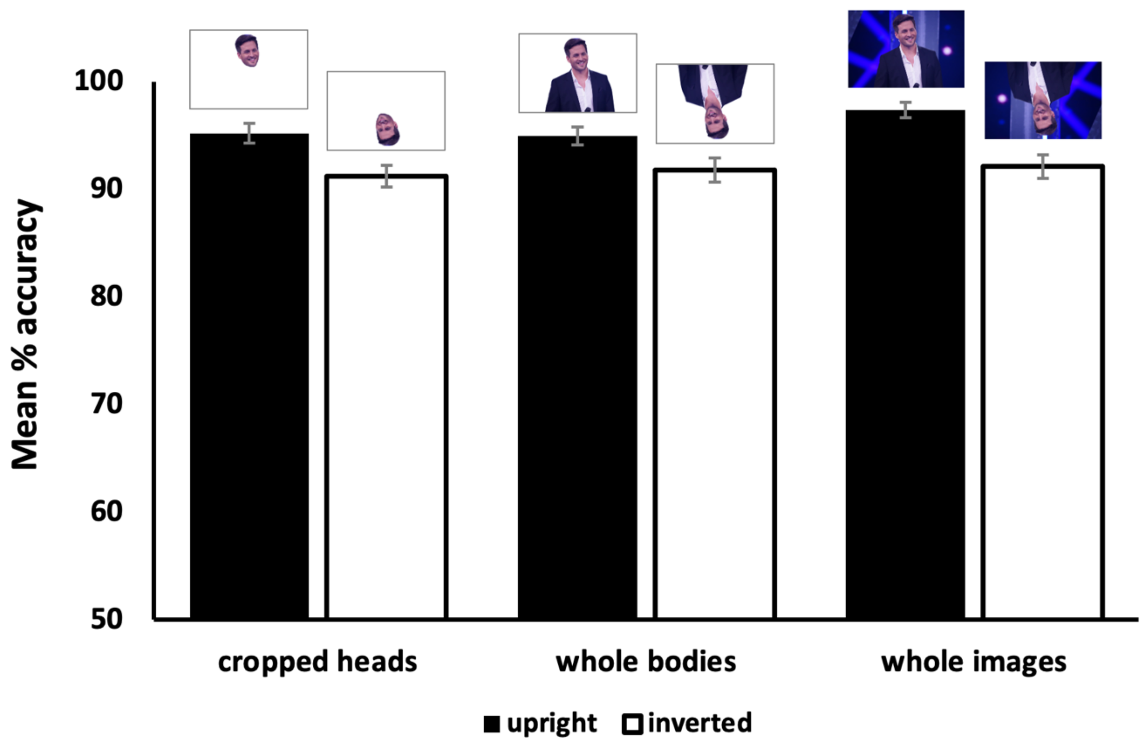



**Figure 3.2.6.** Mean percentage accuracy for recognition of unfamiliar identities shown upright or inverted, and as cropped heads, whole bodies or whole images.

*Unfamiliar Identities: Inter-subject variability*

For the cropped heads, 36% of participants showed an inversion effect, 51% showed no effect, and 13% showed the opposite effect, as shown in Figure 3.2.7. Similarly, for the whole bodies, 33% of participants showed an inversion effect, 49% showed no effect, and 18% showed the opposite effect. Similarly again, for the whole images, 37% of participants showed an inversion effect, 54% showed no effect, and 9% showed the opposite effect.

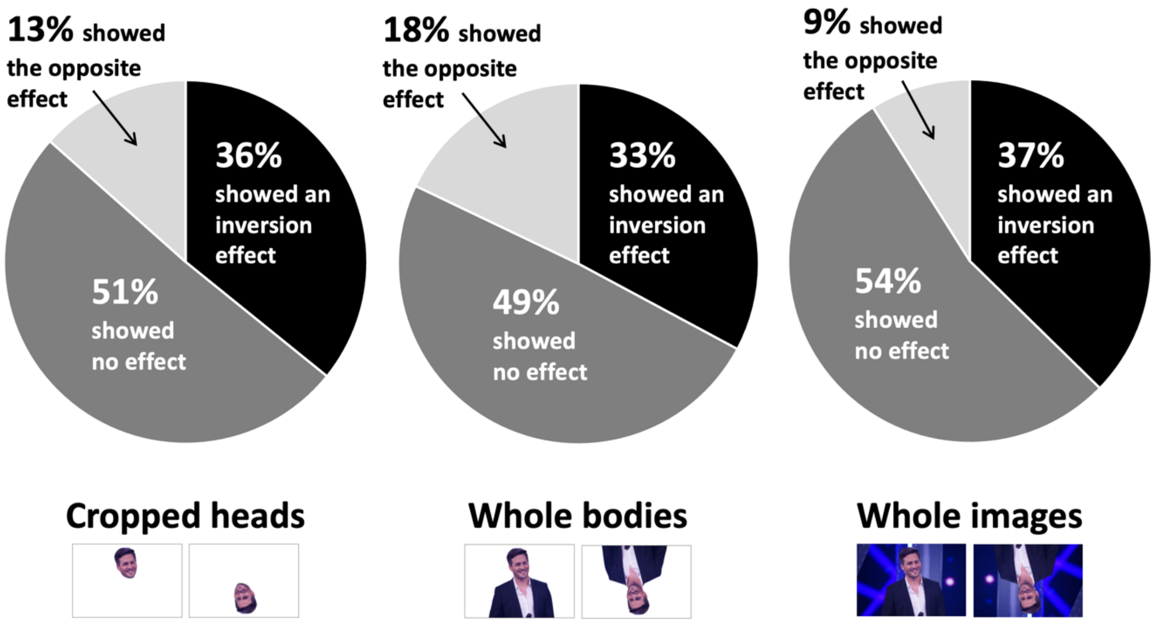



**Figure 3.2.7.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for unfamiliar identities shown as cropped heads, whole bodies or whole images.

*Unfamiliar Identities: RTs*

Figure 3.2.8 shows mean RTs for unfamiliar faces. A 2x3 repeated measures ANOVA (whole image / whole body / cropped head; upright / inverted) was performed on the unfamiliar face RT data. The image type data ($p$ = .165) and interaction ($p$ = .175) met the assumption of sphericity. A significant main effect of orientation was found ($F(1,66) = 53.14$, $p < .001$, $\eta_p^2 = 0.45$), however, there was no significant main effect of image type ($F(2,132) = 0.29$, $p = .746$, $\eta_p^2 = 0.00$). No significant interaction was observed ($F(2,132) = 0.65$, $p = .523$, $\eta_p^2 = 0.01$).

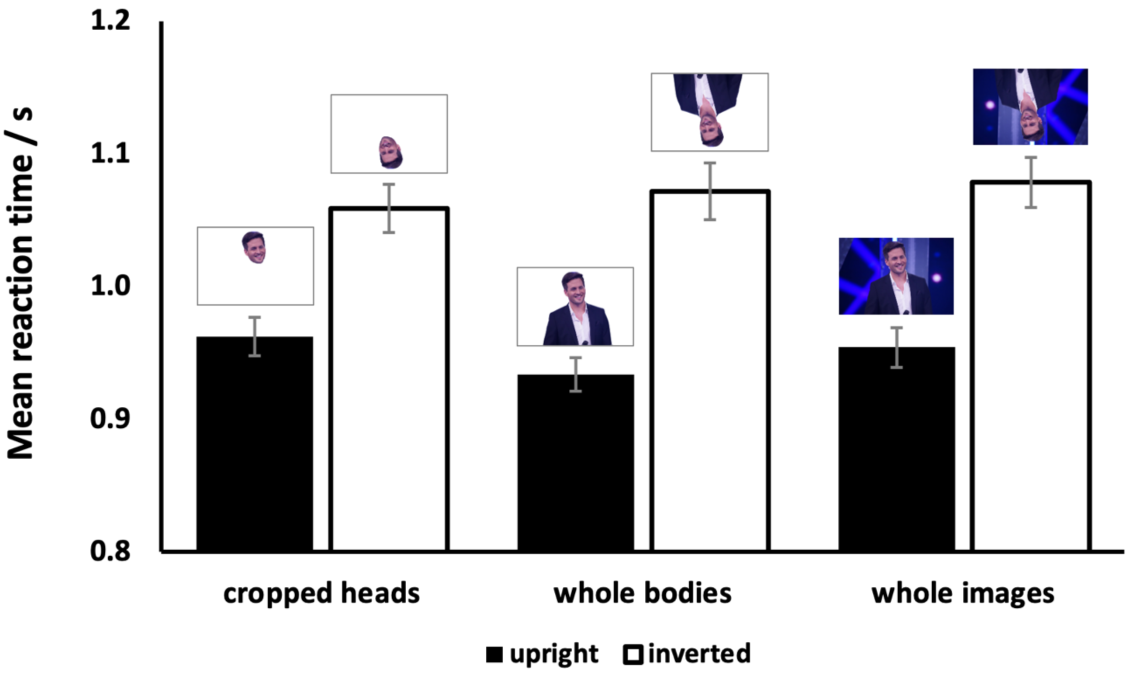



**Figure 3.2.8.** Mean reaction times for recognition of unfamiliar identities shown upright or inverted, and as cropped heads, whole bodies or whole images.

**Discussion**

The results of Experiment 2 show that the cropped faces and whole images replicated the pattern found in Experiment 1: cropped faces suffered much larger inversion effects than whole images, reflected strongly in both the accuracy (27.6% compared to 15.1% decrease) and RT data (0.29s compared to 0.16s increase). However, whole bodies (background removed) and whole images (background included) produced near-identical results to each other in both the accuracy (15.1% compared to 16.3% decrease) and RT data (0.16s increase each). A similar pattern was shown in the inter-subject variability results, with similar proportions of participants showing an inversion effect with the whole bodies and whole images, and more participants showing an inversion effect with the cropped faces.

These results do not support the hypothesis that the presence of a meaningful background cued identity and reduced the inversion effect (Gruppuso et al., 2007). This is shown clearly by the whole bodies condition which had the background removed, yet produced near identical results to the whole images which included the background. When the whole body was shown, the inclusion or removal of the semantically meaningful background had no effect. The large difference in the effect of inversion then lies between the cropped face condition and whole body condition. These results support the hypothesis that the larger inversion effect for cropped faces is due to the artificial nature of the stimuli not representing what the visual system has observed in day to day life (Valentine, 1988). The stimuli in the whole bodies condition are visually plausible, for example in passport photos a person stands in front of a plain wall. However, the stimuli in the cropped faces condition, with the background and body removed and presented inverted, does not resemble what the visual system is used to seeing and may have caused the large decrease in accuracy and increase in RTs.

The results from the first two experiments clearly demonstrate that cropped faces produce large inversion effects, and this can be drastically reduced by using naturalistic whole stimuli. In order to test the reliability of the effect found in Experiments 1 and 2, an attempt to replicate the effect using a different task is presented in Experiment 3.

# Experiment 3: Name-face verification task with familiar faces

## Introduction

Many studies on the face inversion effect use cropped stimuli, however, the results of Experiment 1 and 2 demonstrate that this may be causing an overinflation on the effect of inversion of faces. This represents an unexpected finding, therefore in order to strengthen the claim, the following experiment aimed to replicate the finding using a different task. A name-face verification task (Collin & Byrne, 2010; Ritchie et al., 2018) was used in Experiment 3. It is possible that the pattern of results found in Experiments 1 and 2 may be task-specific and that using a name-face verification task could produce different results, as additional cognitive processes must be performed, e.g. holding a name in working memory. However, if the effect of image cropping is robust and reliable across tasks, then converging evidence from a different task could be found. Based on the results of Experiment 1 and 2, it was hypothesized that cropped faces would demonstrate a large inversion effect and whole images would show a much reduced inversion effect.

## Methods

### *Participants*

78 right-handed participants (53 females, 25 males, mean age = 25.28, SD = 3.46) took part in this experiment. 2 additional participants were excluded from the analysis as they got the response keys the wrong way round. Participants were

recruited from Prolific and were paid £4 to compensate their time. All participants self-reported as being right-handed, currently residing in the United Kingdom, had lived in the UK for at least 10 years, and had normal or corrected-to-normal vision. Informed consent was gained, and the experiment was approved by the University of York Psychology Ethics Board.

***Stimuli***

In this experiment, the 48 famous identities from Experiment 2 were used. The images were again edited to produce 2 versions, a cropped around the head image and a whole image, as shown in Figure 3.3.1. All the names presented in this experiment were the names of the identities used. In the match trials, an identity was paired with its actual name. For the mismatch name-face trials, the names used were the names of the other same-gender celebrities also in the mismatch condition. 15 of the 78 participants experienced a small spelling error in 2 of the famous names. The same 20 images of butterflies from Experiments 1 and 2 were used again in an attention check. (See Appendix E for further examples).

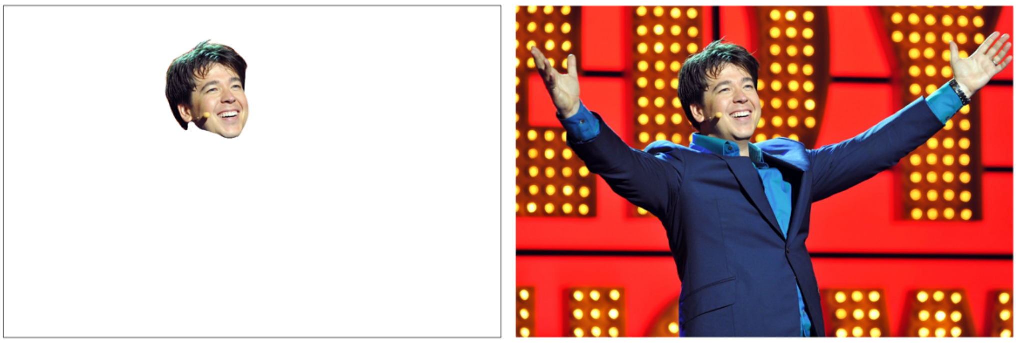


**Figure 3.3.1.** A sample of the stimuli used in Experiment 3. The left photograph demonstrates the cropping around the head and the right photograph shows the whole image.

***Design and Procedure***

Participants saw each identity twice, once cropped around the head and once as a whole image. On one presentation, the face was paired with the correct name, on the other presentation it was paired with the name of a different identity. Identities were counterbalanced across all other conditions.

This experiment had 3 independent variables, each with 2 levels, producing 8 conditions. The image type was either a cropped face or whole image, the trial type was either match or mismatch, and the orientation was either upright or inverted. Participants completed 12 trials in each condition, 96 trials in total. Accuracy and reaction time were recorded.

At the start of a trial, a fixation cross appeared in the centre of the screen for 0.5s on a grey background. A name then appeared in the centre of the screen for 1.5s. The photograph then appeared and remained on screen until the participants made a response, as shown in Figure 3.3.2. Images were again presented in 1.05 x 0.7 ‘height’ units. If a response had not been made after 4.5s, a prompt briefly appeared telling participants to “please answer quickly.” Participants had to indicate via a button press (1 or 0) if the name and photograph were the same person. Again, intermixed were 20 attention check trials where participants pressed “M” when presented with an image of a butterfly. On average, participants were 99.2% accurate. The order of trials was randomised for each participant and the response keys were counterbalanced across participants. A familiarity check again followed the main experiment to remove any unrecognized familiar IDs whereby participants saw the whole upright images again and had to type in their name or identifying information. This resulted in 10.9% of trials being eliminated on average for each participant (1 participant had more than 40% of trials binned, 14 participants had more than 20% of trials binned, 63 participants had below 20% of trials binned). On average, the experiment lasted approximately 22 minutes.

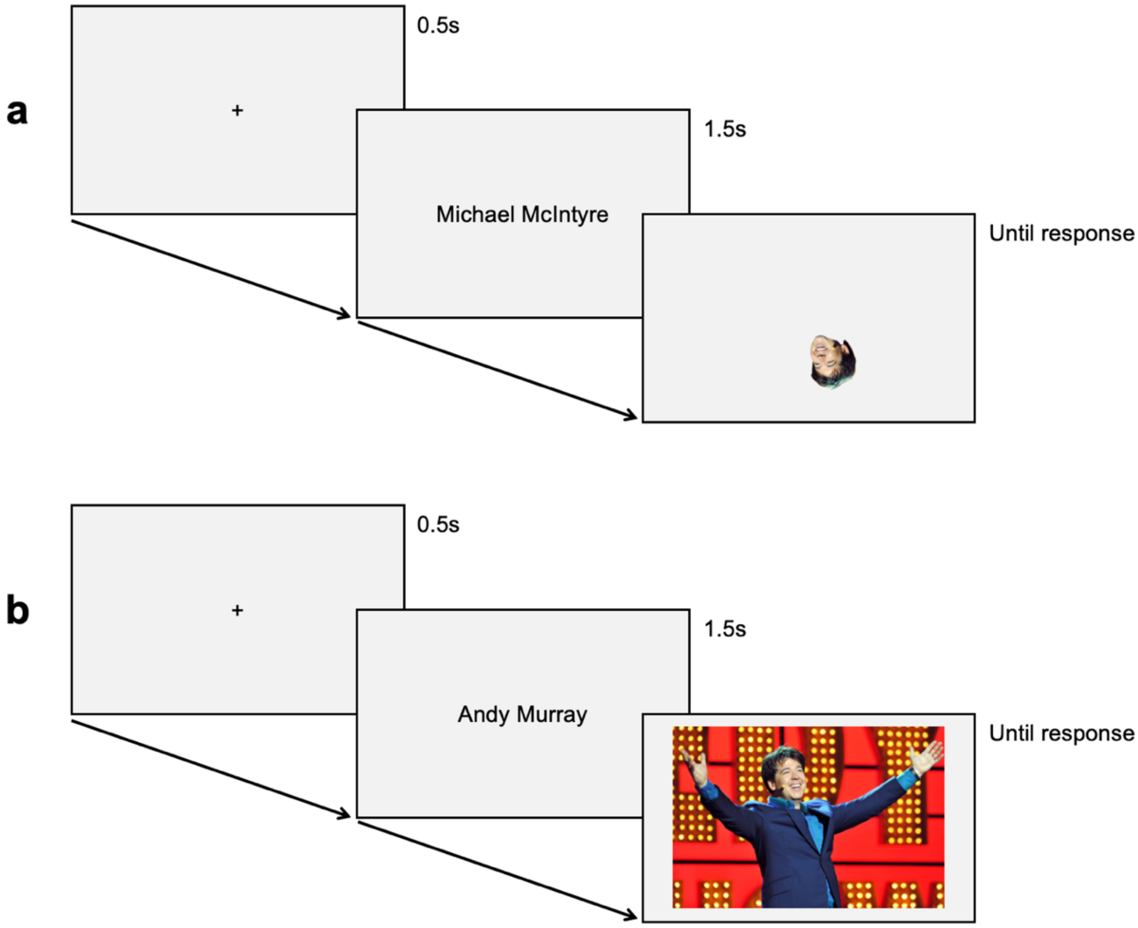



**Figure 3.3.2.** Example trials from Experiment 3. Participants saw a fixation cross followed by a name then a photograph. Participants had to indicate if the name and photograph were of the same person or not. The top row (a) shows a match trial, the bottom row (b) shows a mismatch trial.

**Results**

*Match Trials: Accuracy*

Mean accuracy for match trials is shown in Figure 3.3.3. A 2x2 repeated measures ANOVA (whole / cropped face; upright / inverted) was performed on the match accuracy data. Significant main effects of image type ($F(1,77) = 4.40$, $p = .039$, $\eta_p^2 = 0.05$), and orientation ($F(1,77) = 23.49$, $p < .001$, $\eta_p^2 = 0.23$) were found. No significant interaction was found ($F(1,77) = 1.22$, $p = .274$, $\eta_p^2 = 0.02$).

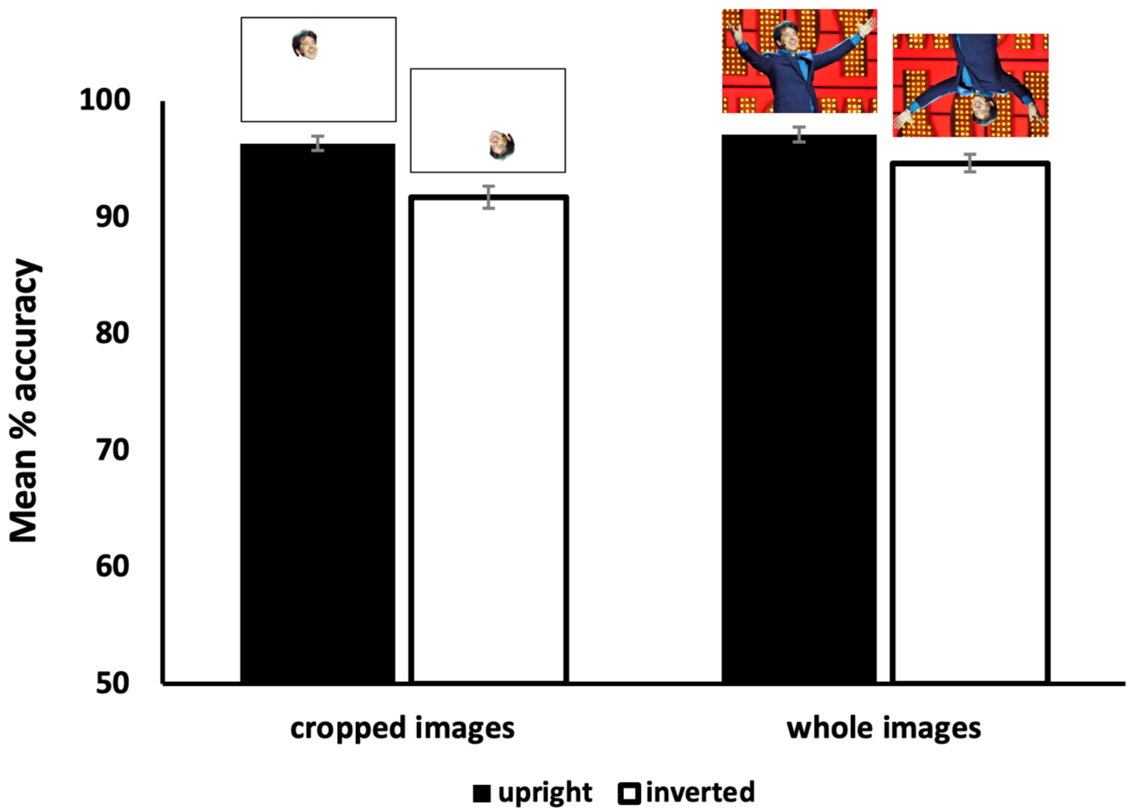



**Figure 3.3.3.** Mean percentage accuracy on match trials for images shown upright or inverted, and as cropped or whole images.

*Match Trials: Inter-subject variability*

For the cropped faces, 44% of participants showed an inversion effect (their accuracy was lower in the inverted condition compared to the upright condition), 40% showed no effect (their accuracy was the same in the upright and inverted conditions), and 17% showed the opposite effect (their accuracy was higher in the inverted condition compared to upright), as shown in Figure 3.3.4. For the whole images, 33% of participants showed an inversion effect, 47% showed no effect, and 19% showed the opposite effect.

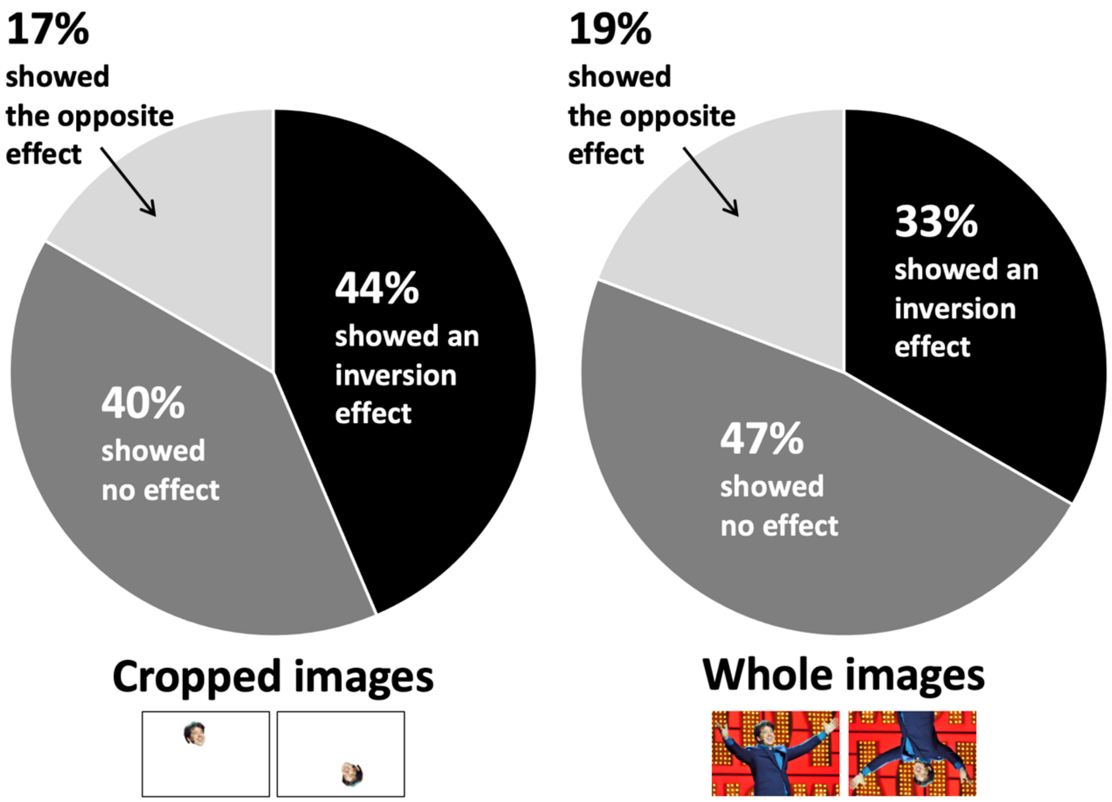



**Figure 3.3.4.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the

opposite effect on match trials for familiar identities shown as cropped or whole images.

*Match Trials: RTs*

Figure 3.3.5 shows mean RTs for match trials. A 2x2 repeated measures ANOVA (whole / cropped image; upright / inverted) was performed on the match RT data. Significant main effects of image type ($F(1,77) = 5.23$, $p = .025$, $\eta_p^2 = 0.06$), and orientation ($F(1,77) = 74.24$, $p < .001$, $\eta_p^2 = 0.49$) were found. No significant interaction was found ($F(1,77) = 0.18$, $p = .672$, $\eta_p^2 = 0.00$).

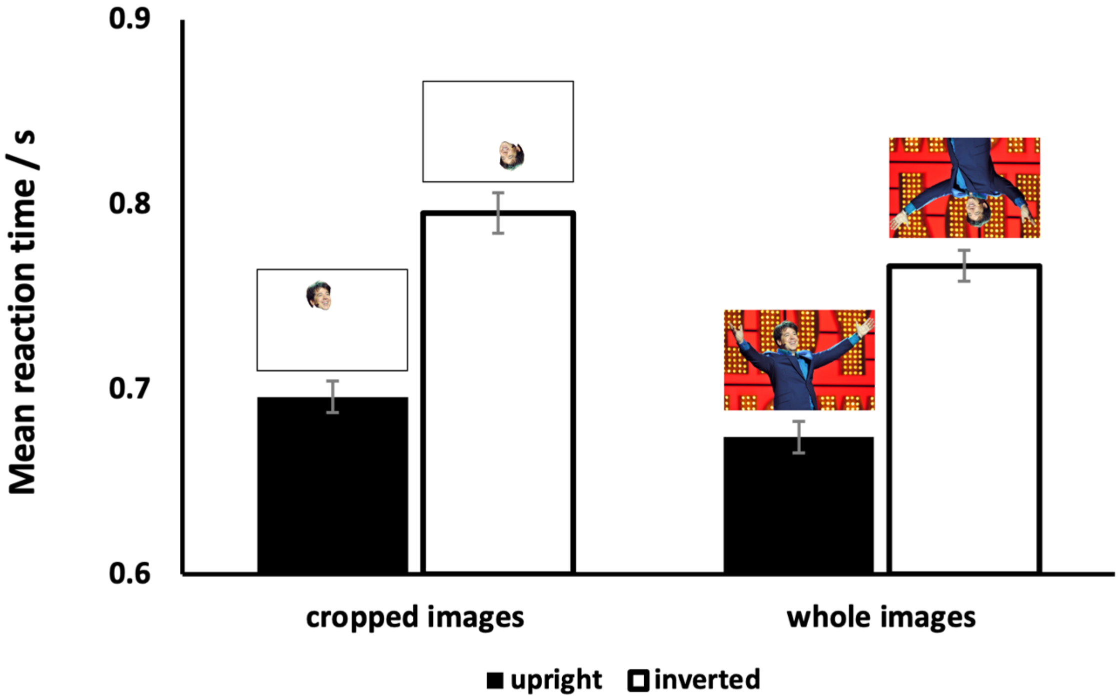



**Figure 3.3.5.** Mean reaction times on match trials for images shown upright or inverted, and as cropped or whole images.

*Mismatch Trials: Accuracy*

Figure 3.3.6 shows mean accuracy for mismatch trials. A 2x2 repeated measures ANOVA (whole / cropped image; upright / inverted) was performed on the mismatch accuracy data. A significant main effect of orientation was found ($F(1,77) = 6.19$, $p = .015$, $\eta_p^2 = 0.07$), however, there was no significant main effect of image type ($F(1,77) = 1.15$, $p = .287$, $\eta_p^2 = 0.02$). No significant interaction was found ($F(1,77) = 2.17$, $p = .144$, $\eta_p^2 = 0.03$).

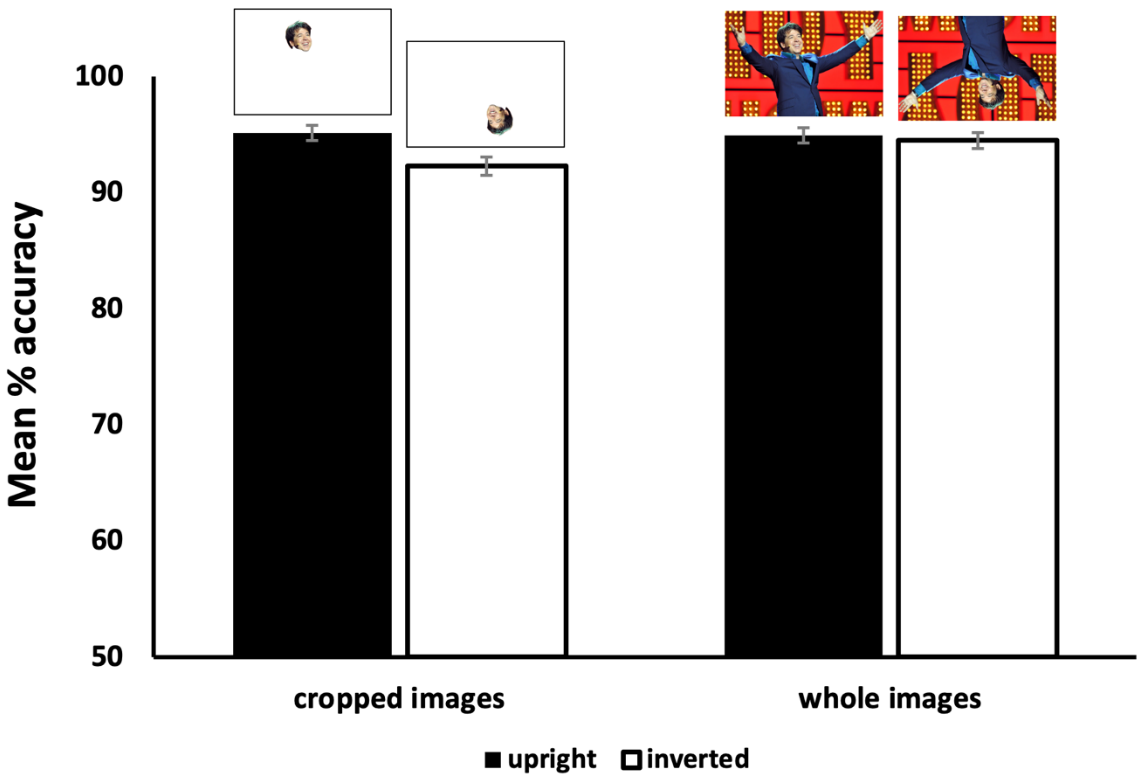



**Figure 3.3.6.** Mean percentage accuracy on mismatch trials for images shown upright or inverted, and as cropped or whole images.

*Mismatch Trials: Inter-subject variability*

For the cropped faces images, 37% of participants showed an inversion effect, 38% showed no effect, and 24% showed the opposite effect, as shown in Figure 3.3.7. For the whole images, 31% of participants showed an inversion effect, 41% showed no effect, and 28% showed the opposite effect.

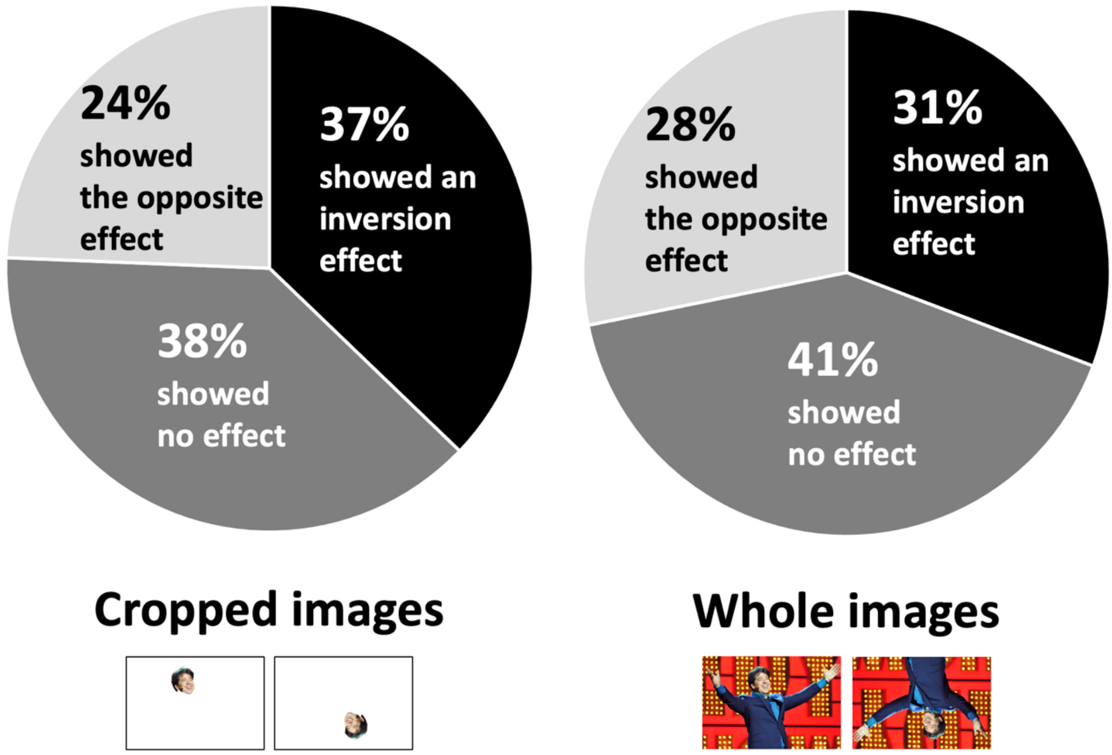



**Figure 3.3.7.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect on mismatch trials for familiar identities shown as cropped or whole images.

*Mismatch Trials: RTs*

Figure 3.3.8 shows mean RTs for mismatch trials. A 2x2 repeated measures ANOVA (whole / cropped image; upright / inverted) was performed on the mismatch RT data. A significant main effect of orientation was found ($F(1,77) = 44.08$, $p < .001$, $\eta_p^2 = 0.36$), however, there was no significant main effect of image type ($F(1,77) = 1.11$, $p = .294$, $\eta_p^2 = 0.01$). No significant interaction was found ($F(1,77) = 0.15$, $p = .697$, $\eta_p^2 = 0.00$).

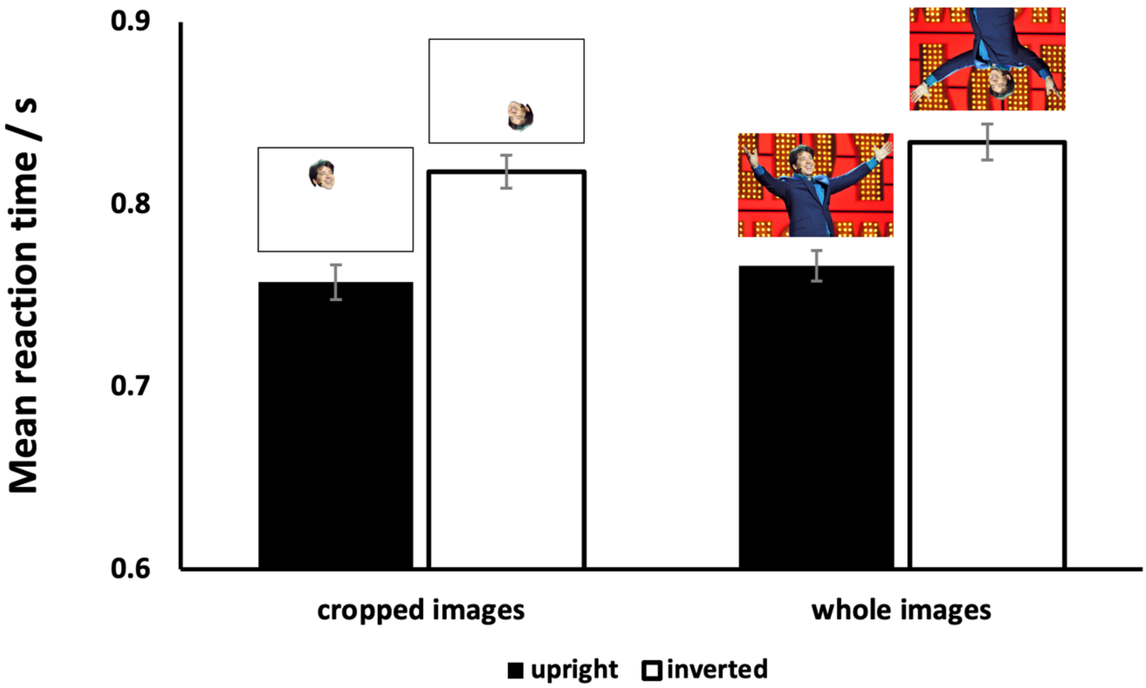



**Figure 3.3.8.** Mean reaction times on mismatch trials for images shown upright or inverted, and as cropped or whole images.

**Discussion**

In this experiment, minimal differences were seen between conditions, with accuracy ranging only between 90-100%. This unfortunately reflects ceiling effects.

This may be due to two reasons. Firstly, the task was very easy. In the match trials, because the correct names preceded the faces, the face representation was primed and the search for the correct face representation was reduced to a search of just one. In combination with the excellent processing of familiar faces, the task was not cognitively demanding for participants, even in the inverted cropped face condition. Similar high accuracy performance has been found in other experiments (Collin & Byrne, 2010; Ritchie et al., 2018).

Secondly, the experiment design contributed to the task being easy. Images of each identity were shown twice, once in the cropped face and once in the whole image conditions. Participants may have been able to detect this and use it to guess the correct answer on the second presentation of the identity. This idea was confirmed by comparing participants accuracy for the first (mean = 92.15%, SD = 6.51) and second (mean = 93.31%, SD = 5.46) presentation of identities using a paired t-test – participants were significantly more accurate at responding to the second presentation of an identity (t (77) = 2.04, *p* = .045). Participants responded equally quickly to the first (mean = 0.76s, SD = 0.15) and second presentation (mean = 0.76s, SD = 0.16), (t (77) = 0.82, *p* = .935).

Whilst the interaction in the match accuracy data failed to reach significance, the same pattern found in the previous experiments is found again here, but was much smaller – the inversion effects for whole images were smaller compared to the cropped faces. The pattern is also found in the inter-subject variability results: fewer participants showed an inversion effect in the whole image condition compared to the cropped face condition. Whilst this experiment does not provide compelling evidence, an interesting point is found by viewing the other side of the coin. The face inversion effect is thought to be very robust, having been replicated many times over, and faces typically produce large inversion effects around 15 - 25% (Robbins & McKone, 2007). However, when using a name-face verification task with familiar faces, only a 2.3 - 4.6% inversion effect size is found in the present high-powered study. Although this study did not provide compelling evidence in support of the hypothesis, the lack of replication using the inverted cropped faces casts some doubt on the reliability of the face inversion effect itself, demonstrating that it is not found across all tasks.

# Experiment 4: Old-new memory task with familiar faces

## Introduction

In the first two experiments, a large inversion effect was found for face stimuli which were harshly cropped, eliminating the body and background. However, the inversion effect was reduced for whole naturalistic images, more akin to inversion effects seen for non-face objects (see Experiment 1). Experiment 3 aimed to replicate this novel effect using a different task, however, any pattern was masked by ceiling effects. Therefore, in this experiment, a further attempt is made to replicate the reduced inversion effect for whole images using a different task. In Yin's (1969) seminal study on the face inversion effect, an old-new memory task was used. In order to make a convincing argument that the face inversion effect is drastically reduced by using naturalistic whole stimuli, Yin's (1969) original task is used here. Participants viewed images of famous faces, which at test were mixed with distractors, and participants indicated which images they had seen before. It was hypothesized that cropped faces would produce a large inversion effect and whole images would produce a small inversion effect.

## Methods

### *Participants*

97 right-handed participants (74 females, 23 males, mean age = 25.41, SD = 3.37) took part in this experiment. 15 additional participants were excluded from the

analysis (10 got the response keys the wrong way round, 1 participant experienced a data recording issue, and 4 participants had missing cases in their data). Participants were recruited from Prolific and were paid £4 to compensate their time. All participants self-reported as being right-handed, currently residing in the United Kingdom, had lived in the UK for at least 10 years, and had normal or corrected-to-normal vision. Informed consent was gained, and the experiment was approved by the University of York Psychology Ethics Board.

***Stimuli***

In this experiment, the 48 famous identities from Experiment 2 and 3 were used. The images were again edited to produce 2 versions, a cropped around the head image and a whole image, see Figure 3.4.1. (See Appendix E for further examples).

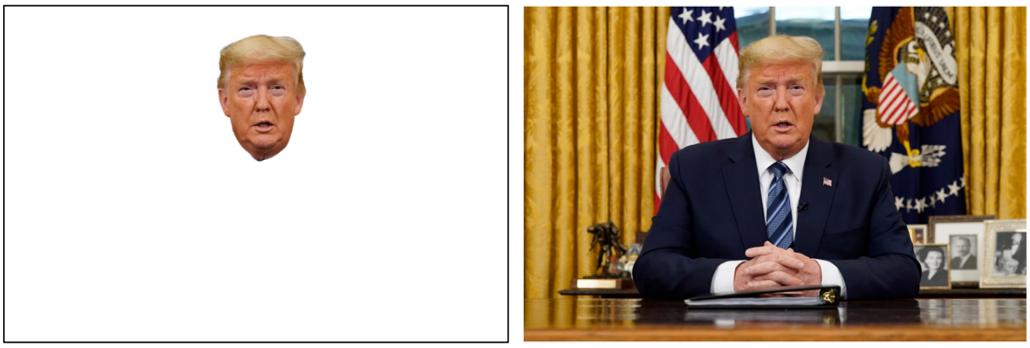


**Figure 3.4.1.** A sample of the stimuli used in Experiment 4. The left photograph demonstrates the cropping around the head and the right photograph shows the whole image.

***Design and Procedure***

In this memory experiment, the identities were split into two sets: half were shown during a learning phase and half were used as distractors in the test phase. During the learning phase, all of the images were presented upright, half were cropped faces and half whole images of identities. The same images were shown at test, either upright or inverted, mixed in with the distractor images.

This experiment had 3 independent variables, each with 2 levels, producing 8 conditions. The image type was either a cropped face or whole image, the orientation at test was either upright or inverted, and the faces were either learned or distractors. Participants completed 6 trials in each condition, 48 trials in total. Accuracy and reaction time were recorded.

During the learning phase, at the start of a trial, a fixation cross appeared in the centre of the screen for 0.5s on a grey background. A face then appeared in the centre of the screen for 1.5s, as shown in Figure 3.4.2. Images were again presented in 1.05 x 0.7 ‘height’ units. The order of trials was randomised for each participant. The learning phase lasted approximately 1 minute.

The test phase immediately followed the learning phase. During the test phase, at the start of a trial, a fixation cross appeared in the centre of the screen for 0.5s on a grey background. The photograph then appeared and remained on screen until the participants made a response. Participants had to indicate via a button press (1 or 0) if they saw the person during the learning phase. If a response had not been made after 4.5s, a prompt briefly appeared telling participants to “please respond quickly.” The order of trials was randomised for each participant and the response keys were counterbalanced across participants. A familiarity check again followed the main experiment to remove any unrecognized familiar IDs, whereby participants saw the whole upright images again and had to type in their name or identifying information. This resulted in 13.8% of trials being eliminated on average for each participant (5 participants had more than 30% of trials binned, 17 participants had more than 20% of trials binned, 73 participants had below 20% of trials binned). On average, the experiment lasted approximately 17 minutes.

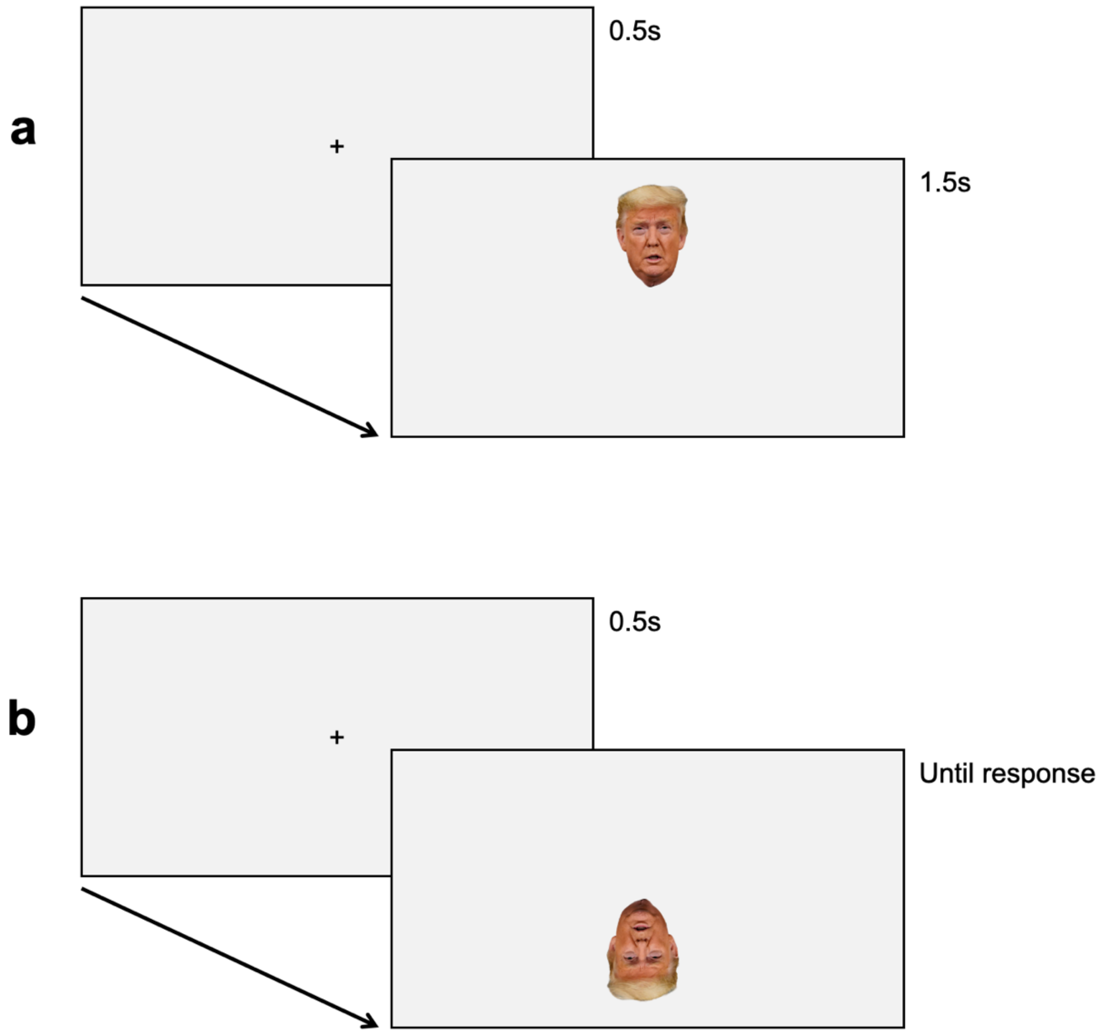



**Figure 3.4.2.** Example trials from Experiment 4. (a) In the learning phase, participants saw a fixation cross followed by an upright photograph. (b) In the test phase, participants saw a fixation cross followed by a photograph and had to indicate whether they saw the person in the learning phase or not.

**Results**

*Learned Identities: Accuracy*

Figure 3.4.3 shows the mean accuracy for learned identities. A 2x2 repeated measures ANOVA (whole / cropped face; upright / inverted) was performed on the learned accuracy data. No significant main effects of image type (F(1,96) = 3.45, $p$ = .066, $\eta_p^2$ = 0.04), nor orientation (F(1,96) = 2.62, $p$ = .109, $\eta_p^2$ = 0.03) were found. No significant interaction was found (F(1,96) = 0.00, $p$ = .986, $\eta_p^2$ = 0.00).

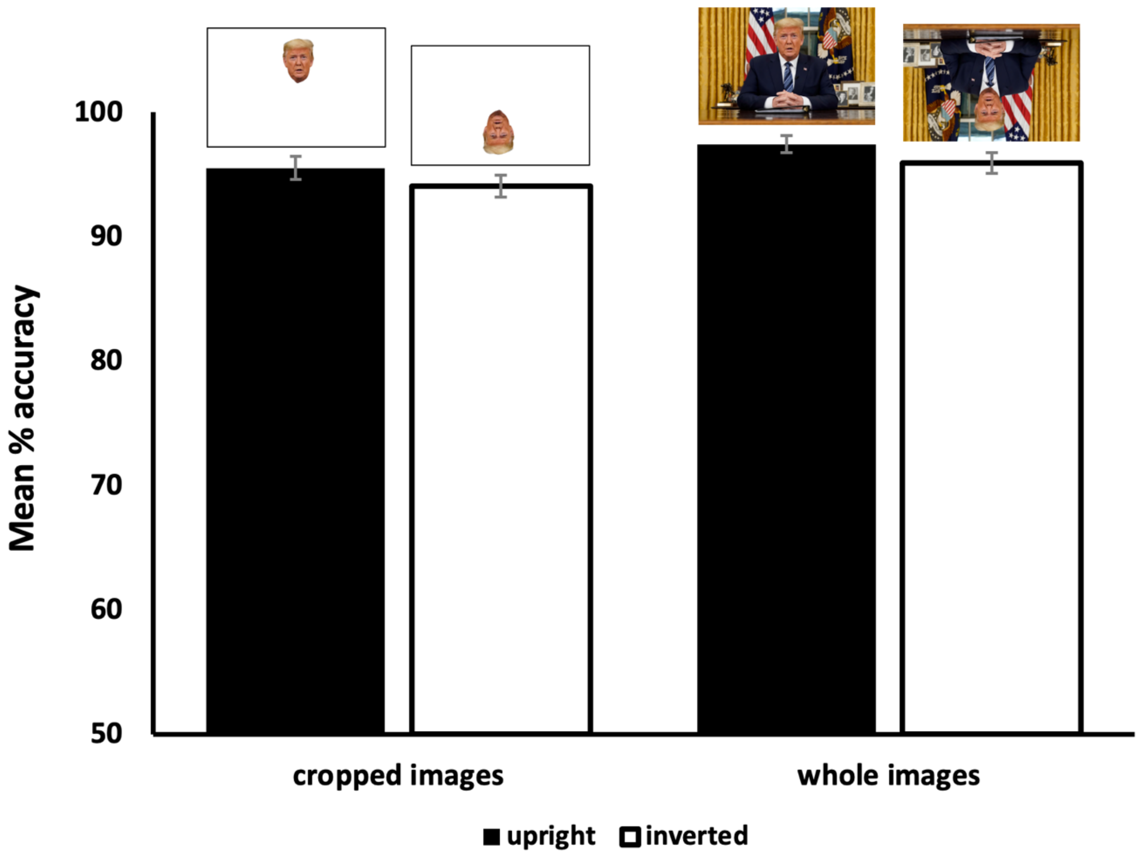



**Figure 3.4.3.** Mean percentage accuracy for learned famous identities shown upright or inverted, and as cropped or whole images.

*Learned Identities: Inter-subject variability*

For the cropped faces, 28% of participants showed an inversion effect (their accuracy was lower in the inverted condition compared to the upright condition), 63% showed no effect (their accuracy was the same in the upright and inverted conditions), and 9% showed the opposite effect (their accuracy was higher in the inverted condition compared to upright) as shown in Figure 3.4.4. For the whole images, 12% of participants showed an inversion effect, 79% showed no effect, and 9% showed the opposite effect.

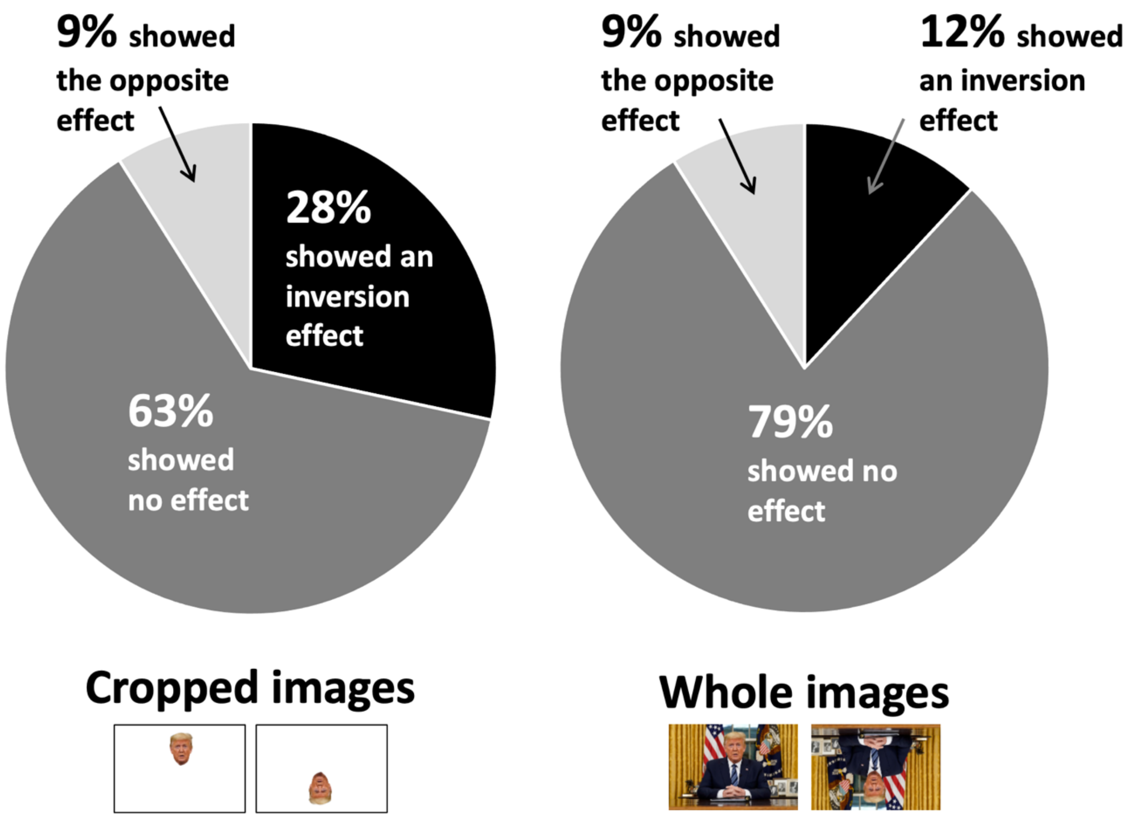



**Figure 3.4.4.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for learned familiar identities shown as cropped or whole images.

*Learned Identities: RTs*

Mean reaction times for learned identities are shown in Figure 3.4.5. A 2x2 repeated measures ANOVA (whole / cropped face; upright / inverted) was performed on the learned RT data. Significant main effects of image type (F(1,96) = 9.83, $p$ = .002, $\eta_p^2$ = 0.09), and orientation (F(1,96) = 95.16, $p$ < .001, $\eta_p^2$ = 0.50) were found. No significant interaction was found (F(1,96) = 0.09, $p$ = .769, $\eta_p^2$ = 0.00).

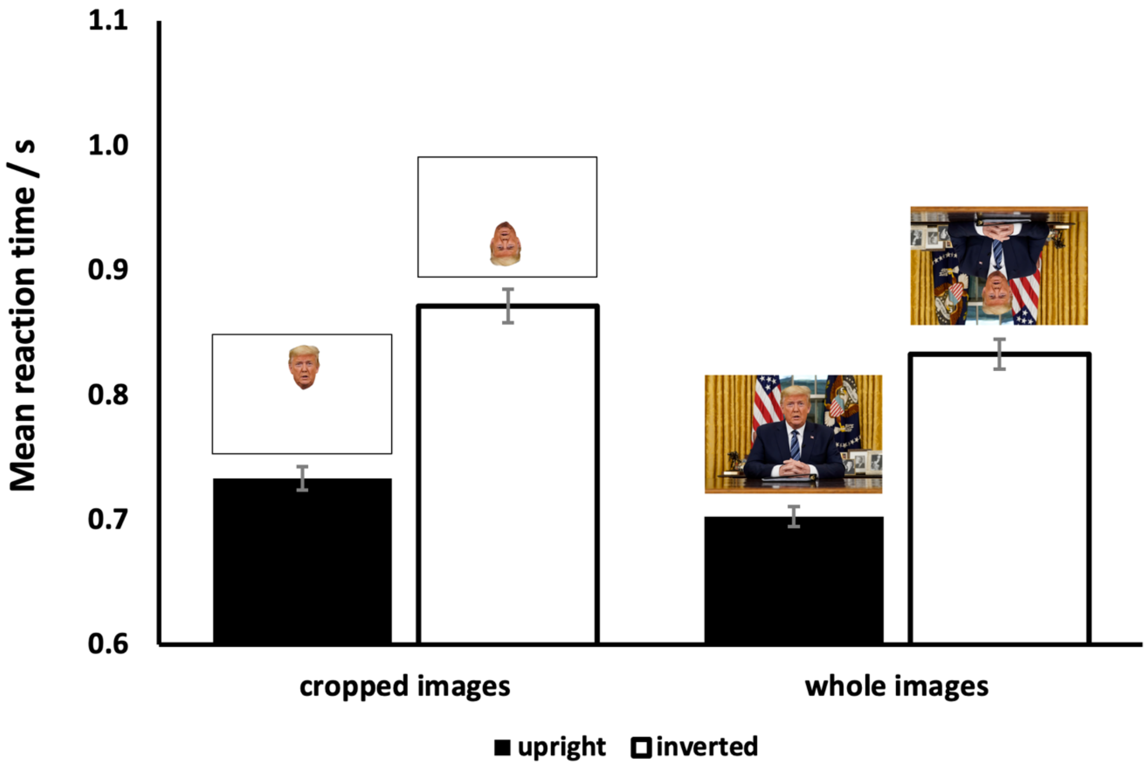



**Figure 3.4.5.** Mean reaction times for learned famous identities shown upright or inverted, and as cropped or whole images.

*Novel Identities: Accuracy*

Mean percentage accuracy for novel identities is shown in Figure 3.4.6. A 2x2 repeated measures ANOVA (whole / cropped face; upright / inverted) was performed on the novel accuracy data. A significant main effect of image type ($F(1,96) = 10.11$, $p = .002$, $\eta_p^2 = 0.10$) was found, but the main effect of orientation was not significant ($F(1,96) = 2.47$, $p = .119$, $\eta_p^2 = 0.03$), nor was the interaction ($F(1,96) = 0.13$, $p = .720$, $\eta_p^2 = 0.00$).

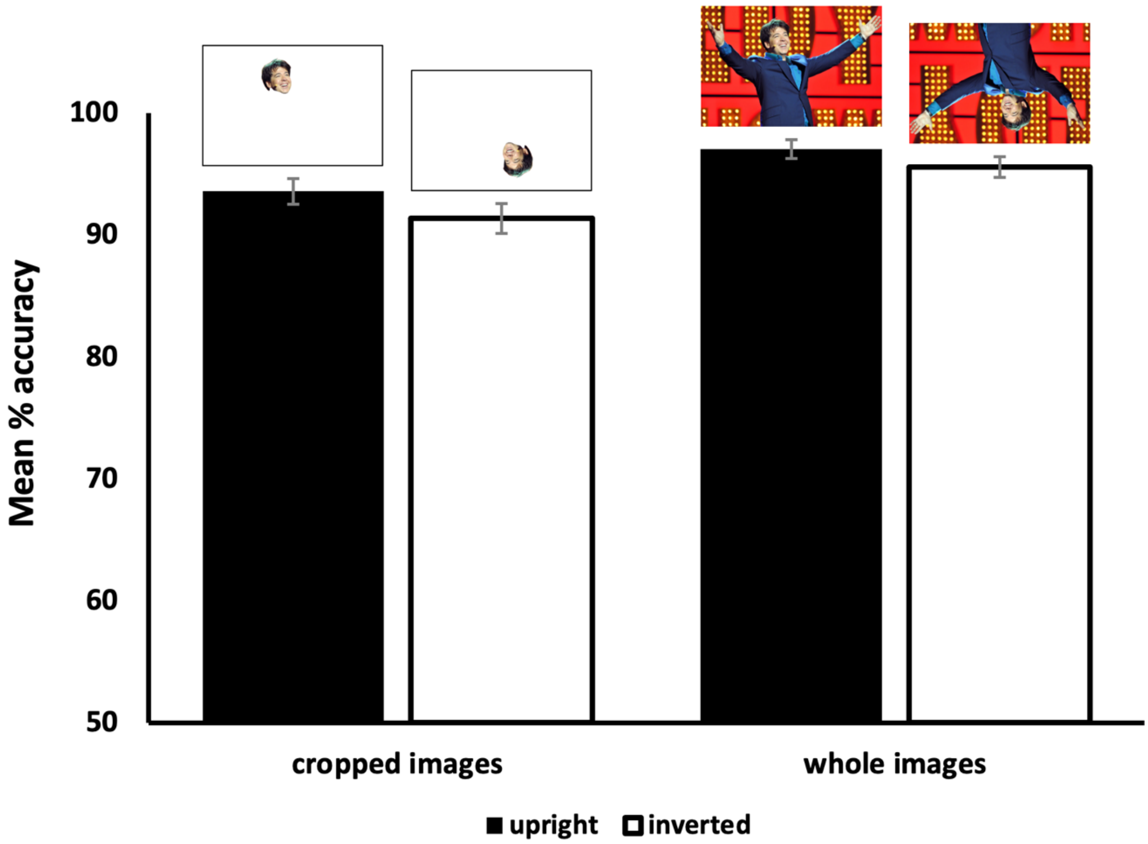



**Figure 3.4.6.** Mean percentage accuracy for novel images of famous identities shown upright or inverted, and as cropped or whole images.

*Novel Identities: Inter-subject variability*

For the cropped faces images, 33% of participants showed an inversion effect, 48% showed no effect, and 19% showed the opposite effect, as shown in Figure 3.4.7.

For the whole images, 15% of participants showed an inversion effect, 73% showed no effect, and 12% showed the opposite effect.

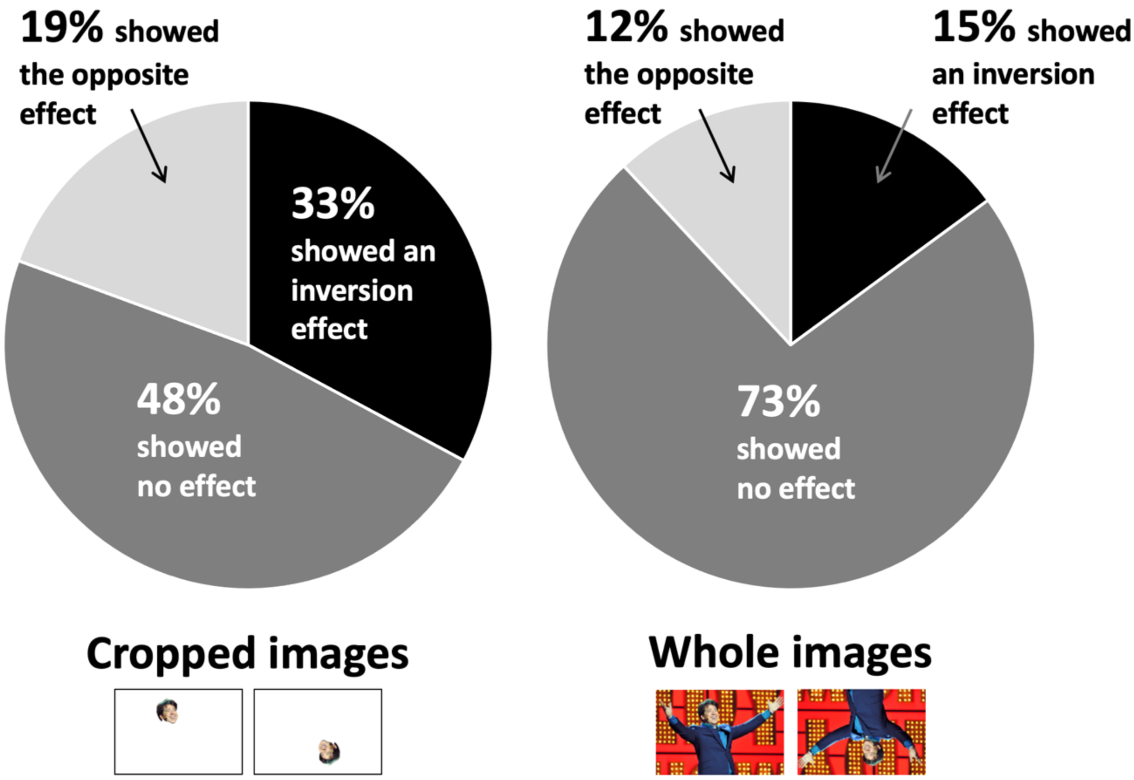



**Figure 3.4.7.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for novel familiar identities shown as cropped or whole images.

*Novel Identities: RTs*

Mean reaction times for novel identities RTs are shown in Figure 3.4.8. A 2x2 repeated measures ANOVA (whole / cropped face; upright / inverted) was performed on the novel RT data. Significant main effects of image type ($F(1,96) = 30.48$, $p < .001$, $\eta_p^2 = 0.24$), and orientation ($F(1,96) = 57.94$, $p < .001$, $\eta_p^2 = 0.38$) were found. A significant interaction was observed ($F(1,96) = 6.06$, $p = .016$, $\eta_p^2 = 0.06$). Simple main effects tests revealed that the interaction reflected the fact that whole image inversion led to a 0.09s increase in RTs ($p < .001$), whereas cropped face inversion led to a 0.16s increase in RTs ($p < .001$).

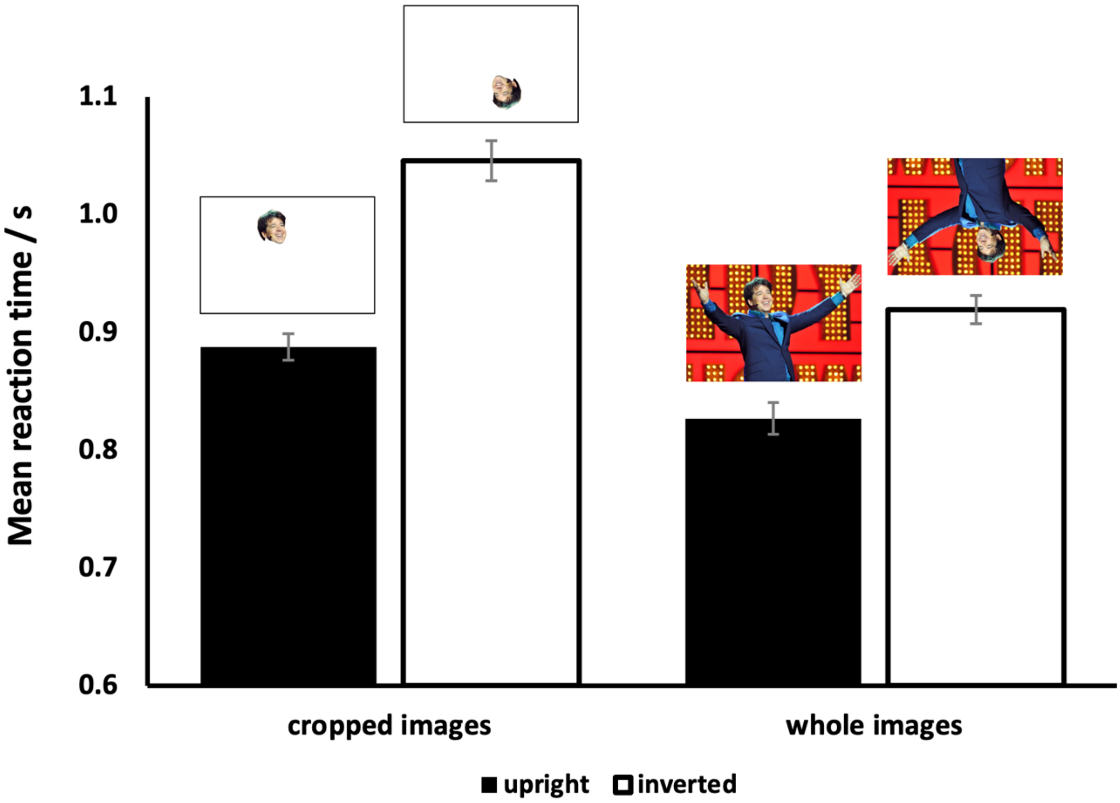



**Figure 3.4.8.** Mean reaction times for novel images of famous identities, shown upright or inverted, and as cropped or whole images.

**Discussion**

In this experiment, minimal differences were found between conditions again, which likely reflects ceiling effects. Although masked by ceiling effects, a slight pattern can be seen in the inter-subject variability data, with fewer participants showing an inversion effect in the whole image condition compared to the cropped face condition as predicted.

The ceiling effects found here may be due to two linked factors: the task being too easy and sub-optimal experimental design. Due to the logistical issues with finding enough famous identities that all participants will recognise, the same 48 famous identities from Experiment 2 and 3 were used. However, as half of the identities were used as distractor items, this produced only 6 trials in each condition. This gave participants minimal room to vary in their responses. In an attempt to increase the power due to the small number of trials, a large sample of 97 participants was recruited, however, this proved ineffective. Furthermore, the task itself and the design of the task was too easy. The task was not cognitively demanding enough due to a combination of the small number of items to remember, the fact that the identities were famous, and the test phase began immediately after the learning phase. Furthermore, the cropped faces were cropped ‘generously’ in comparison to many FIE experiments which crop the face to just include the internal facial features. All of these factors combined to produce a sub-optimally designed and easy experiment resulting in ceiling effects. These issues are addressed in the following experiment.

# Experiment 5: Old-new memory task with unfamiliar faces

## Introduction

The previous experiment aimed to replicate the finding that using whole naturalistic images reduces the effect of inversion using the old-new memory task from Yin's (1969) seminal paper. In this improved experiment, the key issues of Experiment 4 were addressed. Several important changes were made to improve both the difficulty of the task and the experimental design: unfamiliar faces were used instead of famous identities, faces were cropped more harshly into ovals including mainly the internal features, the number of identities and therefore trials was increased, the sample size was increased, an attention task was included, and an unrelated task was incorporated between the learning and test phases to serve as a delay. Based on Experiment 1 and 2, it was hypothesized that cropped faces would produce larger inversion effects than whole images.

## Methods

### *Participants*

106 right-handed participants (62 females, 44 males, 1 participant identified as "other", mean age = 30.28, SD = 5.56) took part in this experiment. 14 additional participants were excluded from the analysis (12 got the response keys the wrong way round, 1 participant took a 40 minute break resulting in a much longer gap between exposure and test than the other participants, and 1 participant had missing cases).

Participants were recruited from Prolific and were paid £3.50 to compensate their time. All participants self-reported as being right-handed, currently residing in the United Kingdom, had lived in the UK for at least 10 years, and had normal or corrected-to-normal vision. Informed consent was gained, and the experiment was approved by the University of York Psychology Ethics Board.

### *Stimuli*

In this experiment, 96 images of unfamiliar identities were collected, 1 photograph per ID. These were new identities and images not used in previous experiments. All images were cropped to a 1:1 ratio and were in full colour. In the whole image condition, the face photographs included the head and upper body, and the background was clearly visible. In the cropped image condition, the faces were cropped to ovals displaying the chin and internal features, as shown in Figure 3.5.1. In Experiment 1, it was not possible to crop the external features out as the faces were being compared to buildings and an equivalent crop could not be applied to buildings. However, as that was not a constraint in this experiment, a harsher crop more akin to the stimuli used in many FIE experiments could be used.

Each image was edited to create 4 versions: cropped upright, cropped inverted, full image upright, full image inverted. The images were fully counterbalanced across participants, so each participant only saw each identity once in only 1 of the 4 conditions. (See Appendix F for further examples).

**Figure 3.5.1.** A sample of the stimuli used in Experiment 5. The left photograph demonstrates the cropping around the internal features and the right photograph shows the whole image.

***Design and Procedure***

This experiment had the same design as that of Experiment 4: half of the IDs were learnt upright during an exposure phase and half were used as novel distractors at test. At test, half of the images were inverted and half were upright, and half of them were full images and half were cropped ovals.

As this experiment had the same design as Experiment 4, it also had 3 independent variables, each with 2 levels, producing 8 conditions. The image type was either a cropped face or whole image of an identity, the orientation at test was either upright or inverted, and the faces were either learned or distractors. Participants completed 12 trials in each condition, 96 trials in total. Accuracy and reaction time were recorded.

During the learning phase, at the start of a trial, a fixation cross appeared in the centre of the screen for 0.5s on a grey background. A photograph of an identity then appeared in the centre of the screen for 1.5s, as shown in Figure 3.5.2. As an attention check, participants had to indicate if the face was female or male by pressing the “F” or “M” key; participants were on average 97.4% accurate. When the participants made a response, the message “response recorded” appeared at the top of the screen for the remainder of the trial. Participants were told to try to remember the faces as they would be tested on them later. Images were presented in 0.6 x 0.6 ‘height’ units. The order of trials was randomised for each participant. The learning phase lasted approximately 2 minutes.

Participants then completed the classic Stroop task (Stroop, 1935), with the added condition of inversion. 4 colour words (red, yellow, blue and green) were used in the task. Half of the trials were congruent (the word-meaning and the ink colour were the same, e.g. the word “red” displayed in red) and half were incongruent (the word-meaning and ink colour were different, e.g. the word “red” displayed in blue). Half of the trials were presented upright, and half were inverted. This produced 4 conditions, congruent upright, congruent inverted, incongruent upright and incongruent inverted. There were 60 trials in each condition, producing 240 trials in total. The order of trials was randomised for each participant and the response keys were kept constant for all participants. At the start of the trial, a fixation cross appeared in the centre of the screen for 0.5s on a grey background. A word then appeared in the centre of the screen and participants had to indicate the colour of the ink by pressing one of the 4 arrow keys while ignoring the meaning of the word. The task was self-paced. On average, this task lasted approximately 5 minutes.

During the memory test phase, at the start of a trial, a fixation cross appeared in the centre of the screen for 0.5s on a grey background. Next, a photograph appeared and remained on screen until the participants made a response. Participants had to indicate via a button press (1 or 0) if they saw the person during the learning phase or not. If a response had not been made after 4.5s, a prompt briefly appeared telling participants to “please respond quickly.” The order of trials was randomised for each participant and the response keys were counterbalanced across participants. On average, this task lasted approximately 2 minutes and the entire experiment lasted approximately 15 minutes.

## Results

### *Stroop Task: Data Removal*

Data from 5 participants were removed for the Stroop analysis (n = 101). This is because their data was at ceiling in the congruent conditions and at floor (below chance) on the incongruent conditions. Inspection of the raw data suggests participants misunderstood the task and were identifying the word meaning and not the word ink colour. As they still completed the task, their data has not been removed from the main experimental data analysis.

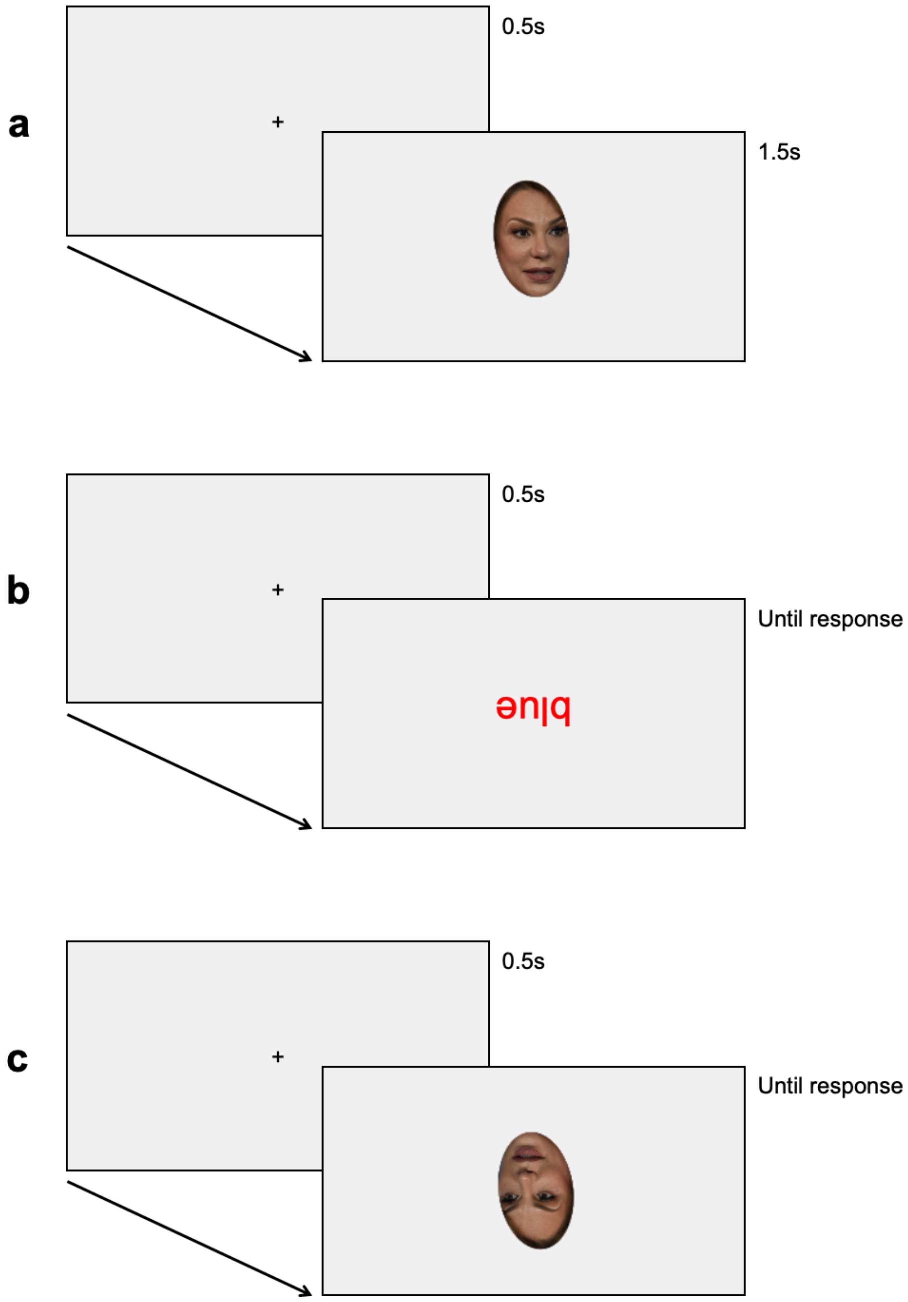



**Figure 3.5.2.** Example trials from Experiment 5. (a) In the learning phase, participants saw a fixation cross followed by an upright photograph. (b) In the Stroop task, participants indicated the colour in which a word was presented in while ignoring the meaning of the word. (c) In the test phase, participants saw a fixation cross followed by a photograph and had to indicate if they saw the person in the learning phase or not.

*Stroop Task: Accuracy*

Figure 3.5.3 shows the mean accuracy. A 2x2 repeated measures ANOVA (congruent / incongruent; upright / inverted) was performed on the Stroop accuracy data. A significant main effect of congruency was found ($F(1,100) = 40.13$, $p < .001$, $\eta_p^2 = 0.29$), however, there was no significant main effect of orientation ($F(1,100) = 2.75$, $p = .101$, $\eta_p^2 = 0.03$). A significant interaction was found ($F(1,100) = 4.25$, $p = .042$, $\eta_p^2 = 0.04$). Simple main effects tests revealed that the interaction reflected the fact that inversion of congruent trials decreased accuracy by 0.4% ($p = .324$) however, inversion of incongruent trials increased accuracy by 1.3% ($p = .018$).

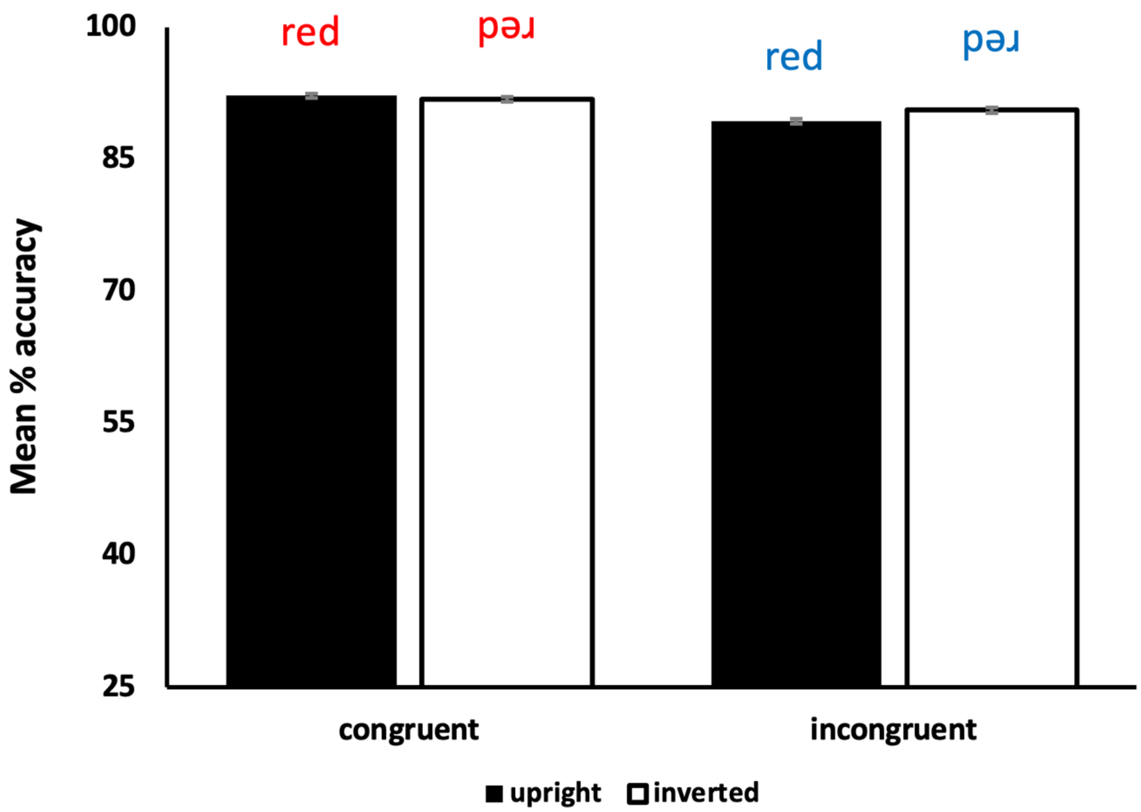



**Figure 3.5.3.** Mean percentage accuracy for upright and inverted, congruent and incongruent trials in the Stroop task.

*Stroop Task: RTs*

Figure 3.5.4 shows the mean reaction times. A 2x2 repeated measures ANOVA (congruent / incongruent; upright / inverted) was performed on the Stroop RT data. A significant main effect of congruency was found ($F(1,100) = 238.07$, $p < .001$, $\eta_p^2 = 0.70$), however, there was no significant main effect of orientation ($F(1, 100) = 1.64$, $p = .203$, $\eta_p^2 = 0.02$). A significant interaction was found ($F(1, 100) = 54.77$, $p < .001$, $\eta_p^2 = 0.35$). Simple main effects tests revealed that the interaction reflected the fact that inversion of congruent trials increased RTs by 0.04s ($p < .001$) and inversion of incongruent trials decreased RTs by 0.06s ($p < .001$).

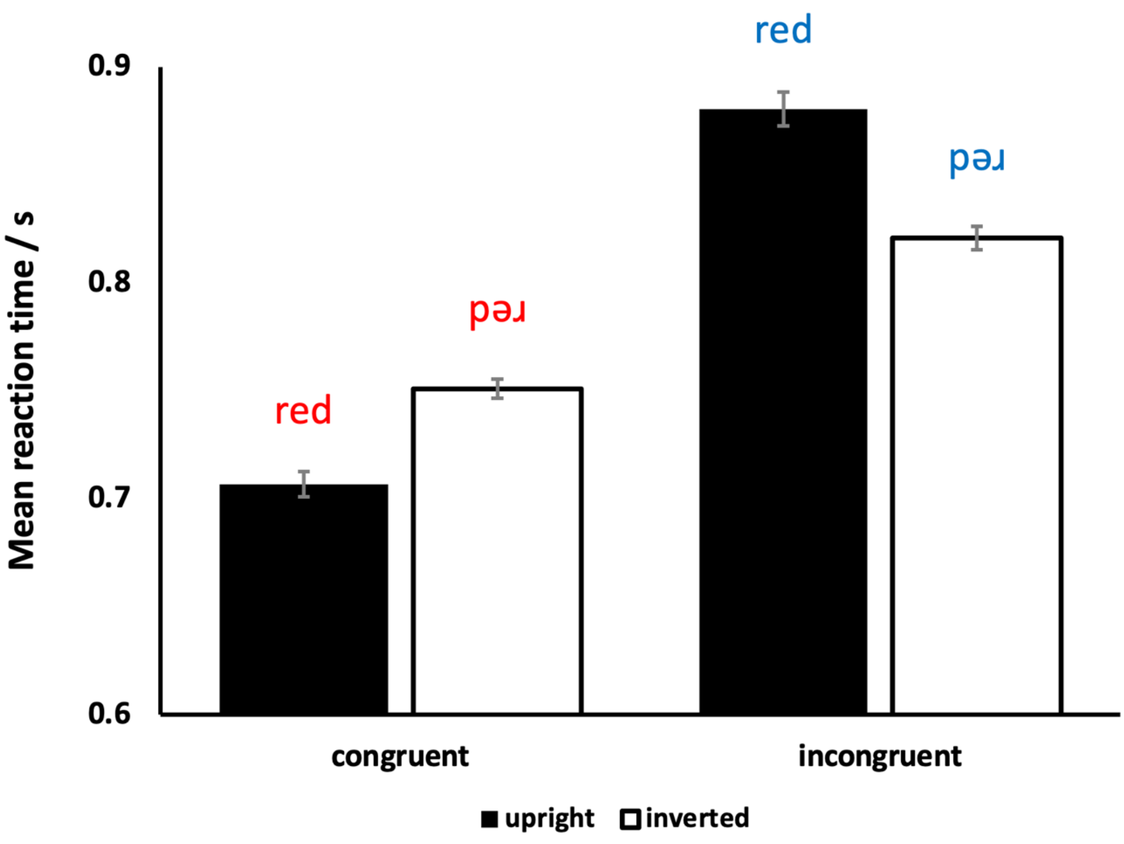



**Figure 3.5.4.** Mean reaction times for upright and inverted, congruent and incongruent trials in the Stroop task.

*Learned Identities: Accuracy*

Figure 3.5.5 shows the mean accuracy for learned identities. A 2x2 repeated measures ANOVA (whole / oval cropped faces; upright / inverted) was performed on the learned accuracy data. A significant main effect of orientation was found ($F(1,105) = 80.80$, $p < .001$, $\eta_p^2 = 0.44$), however, the main effect of image type was not significant ($F(1,105) = 3.58$, $p = .061$, $\eta_p^2 = 0.03$) and the interaction was not significant ($F(1,105) = 0.00$, $p = .975$, $\eta_p^2 = 0.00$).

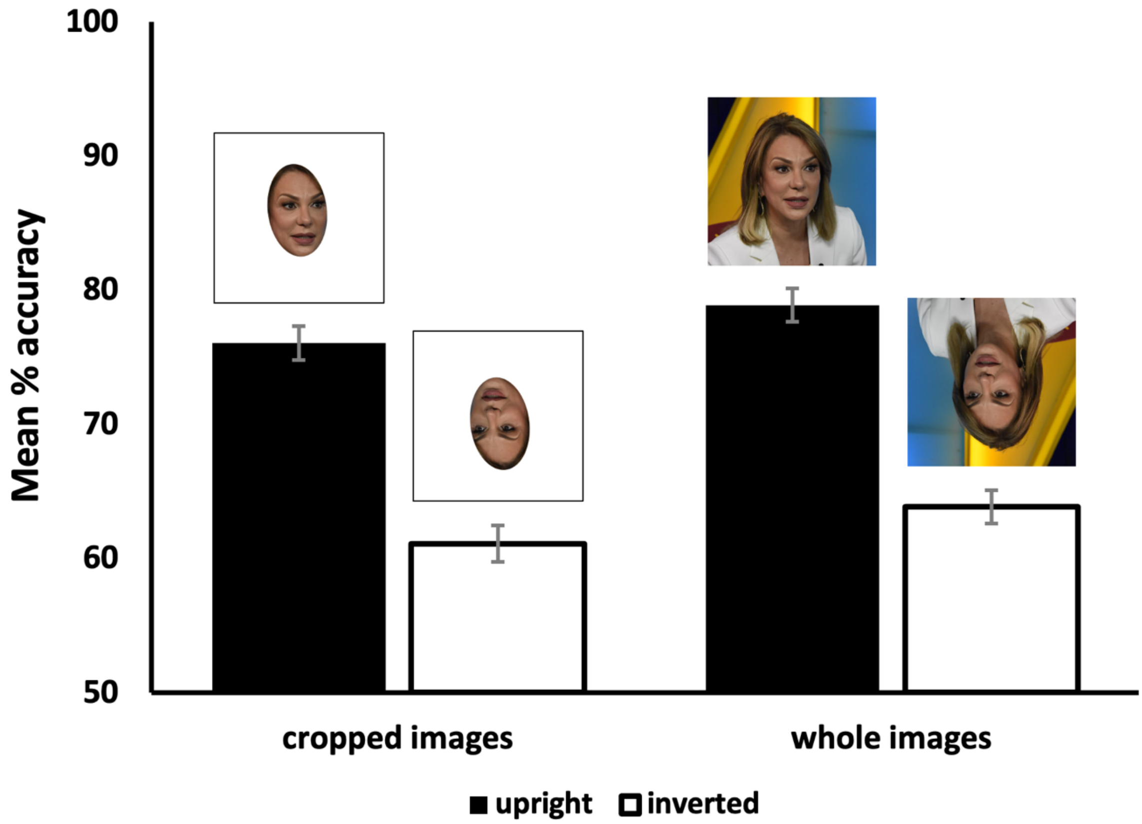



**Figure 3.5.5.** Mean percentage accuracy for learned unfamiliar identities shown upright or inverted, and as cropped ovals or whole images.

*Learned Identities: Inter-subject variability*

For the oval cropped faces, 72% of participants showed an inversion effect (their accuracy was lower in the inverted condition compared to the upright condition), 9% showed no effect (their accuracy was the same in the upright and inverted conditions), and 19% showed the opposite effect (their accuracy was higher in the inverted condition compared to upright), as shown in Figure 3.5.6. Similarly, for the whole images, 68% of participants showed an inversion effect, 12% showed no effect, and 20% showed the opposite effect.

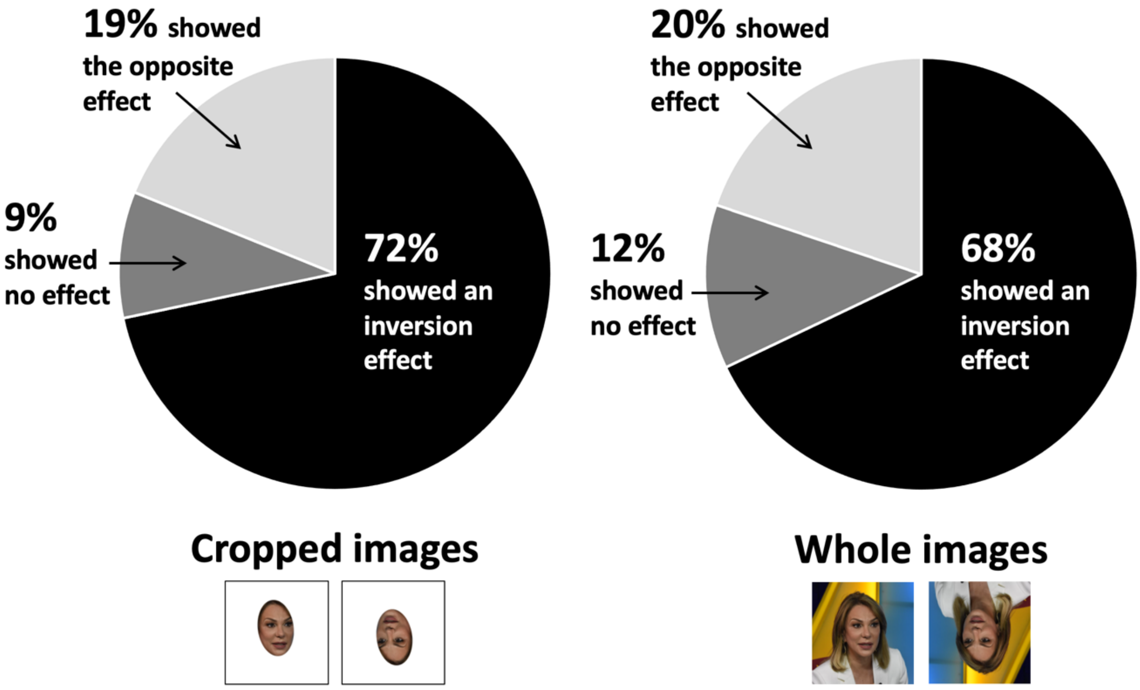



**Figure 3.5.6.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for learned unfamiliar identities shown as cropped ovals or whole images.

*Learned Identities: RTs*

Figure 3.5.7 shows the mean reaction times for learned identities. A 2x2 repeated measures ANOVA (whole image / oval cropped faces; upright / inverted) was performed on the learned RT data. Significant main effects of image type (F(1,105) = 5.41, $p$ = .022, $\eta_p^2$ = 0.05), and orientation (F(1,105) = 62.20, $p$ < .001, $\eta_p^2$ = 0.37) were found. No significant interaction was found (F(1,105) = 0.11, $p$ = .737, $\eta_p^2$ = 0.00).

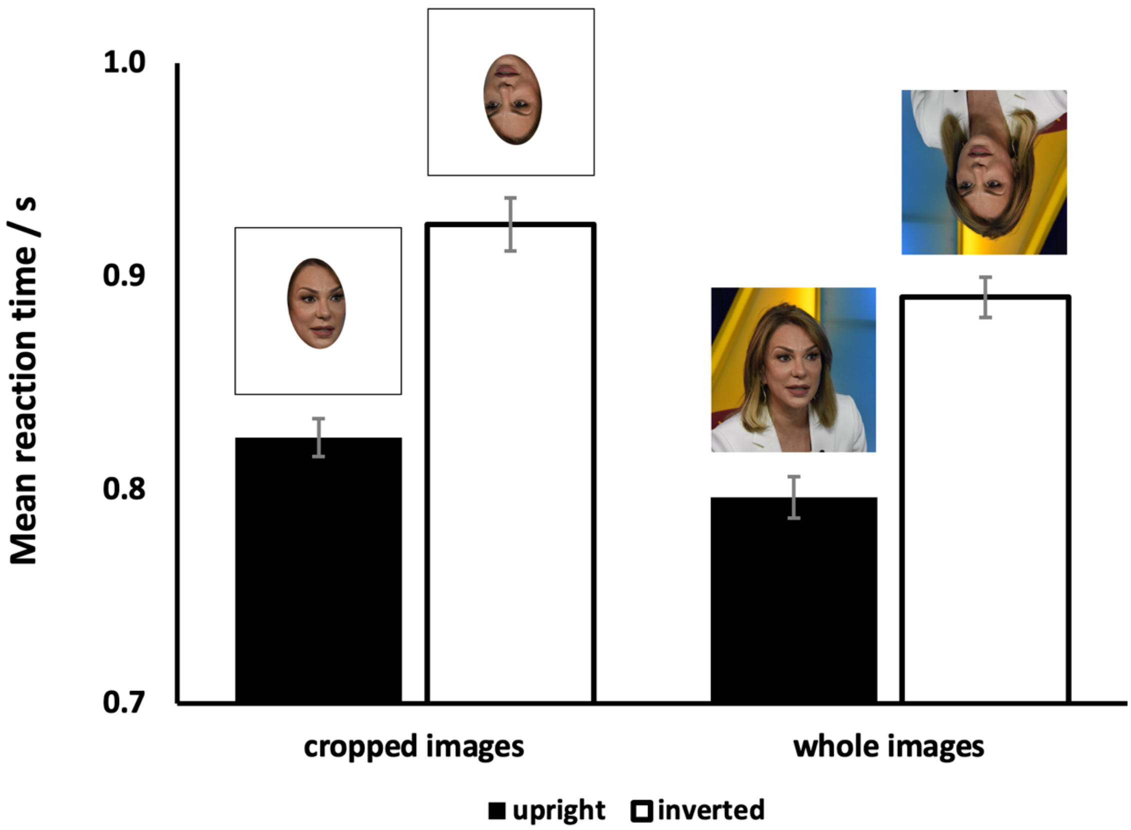



**Figure 3.5.7.** Mean reaction times for learned unfamiliar identities shown upright or inverted, and as cropped ovals or whole images.

*Novel Identities: Accuracy*

Figure 3.5.8 shows the mean accuracy for novel identities. A 2x2 repeated measures ANOVA (whole image / oval cropped faces; upright / inverted) was performed on the novel accuracy data. Significant main effects of image type ($F(1,105) = 93.59$, $p < .001$, $\eta_p^2 = 0.47$), and orientation ($F(1,105) = 18.32$, $p < .001$, $\eta_p^2 = 0.15$) were found. No significant interaction was observed ($F(1,105) = 0.00$, $p = .973$, $\eta_p^2 = 0.00$).

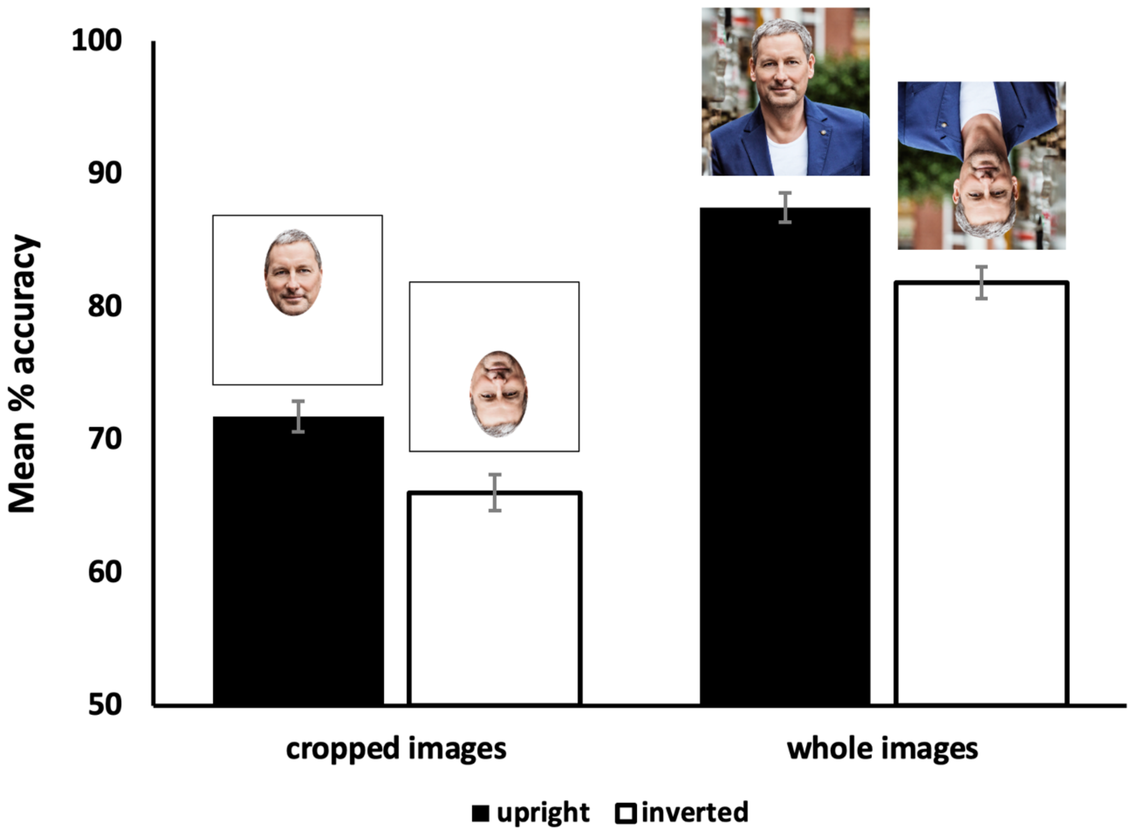



**Figure 3.5.8.** Mean percentage accuracy for novel unfamiliar identities shown upright or inverted, and as cropped ovals or whole images.

*Novel Identities: Inter-subject variability*

For the oval cropped faces, 54% of participants showed an inversion effect, 14% showed no effect, and 32% showed the opposite effect, as shown in Figure 3.5.9. For the whole images, 45% of participants showed an inversion effect, 30% showed no effect, and 25% showed the opposite effect.

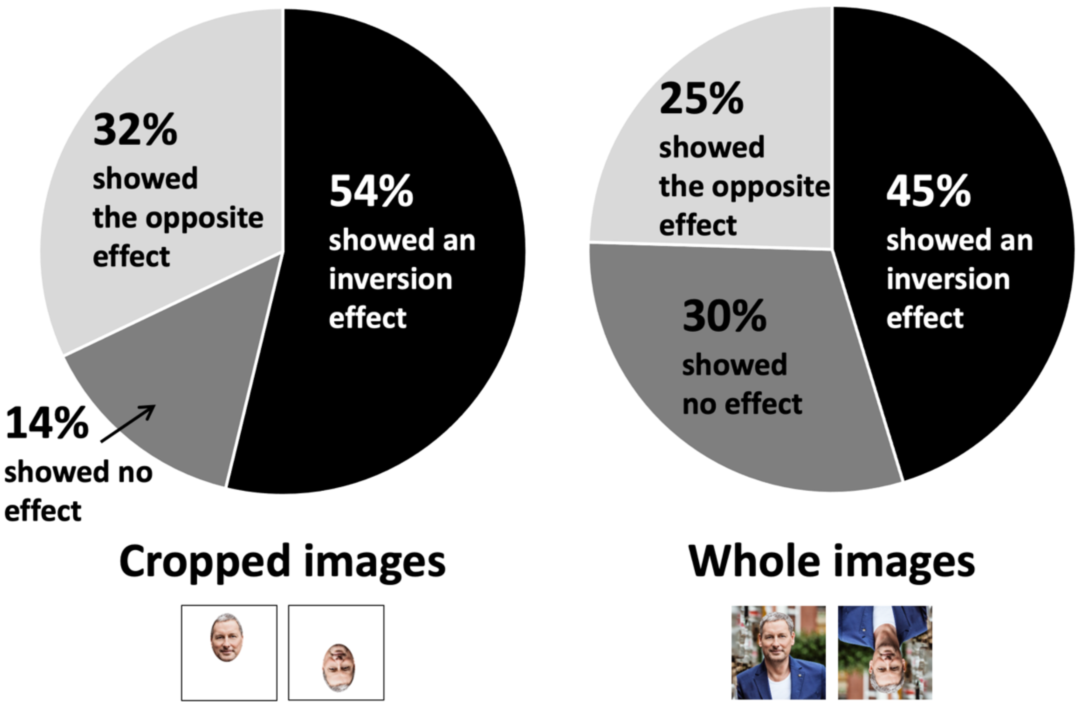



**Figure 3.5.9.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for novel unfamiliar identities shown as cropped ovals or whole images.

*Novel Identities: RTs*

Figure 3.5.10 shows mean reaction times for novel identities. A 2x2 repeated measures ANOVA (whole image / oval cropped faces; upright / inverted) was performed on the novel RT data. A significant main effect of image type ($F(1,105) = 12.55$, $p = .001$, $\eta_p^2 = 0.11$) was found. The main effect of orientation was not significant ($F(1,105) = 3.72$, $p = .057$, $\eta_p^2 = 0.03$). A significant interaction was also found ($F(1,105) = 4.20$, $p = .043$, $\eta_p^2 = 0.04$). Simple main effects tests revealed that the interaction reflected the fact that inversion of cropped images increased RTs by 0.01s ($p = .860$), whereas inversion of whole images increased RTs by 0.08s ($p < .001$).

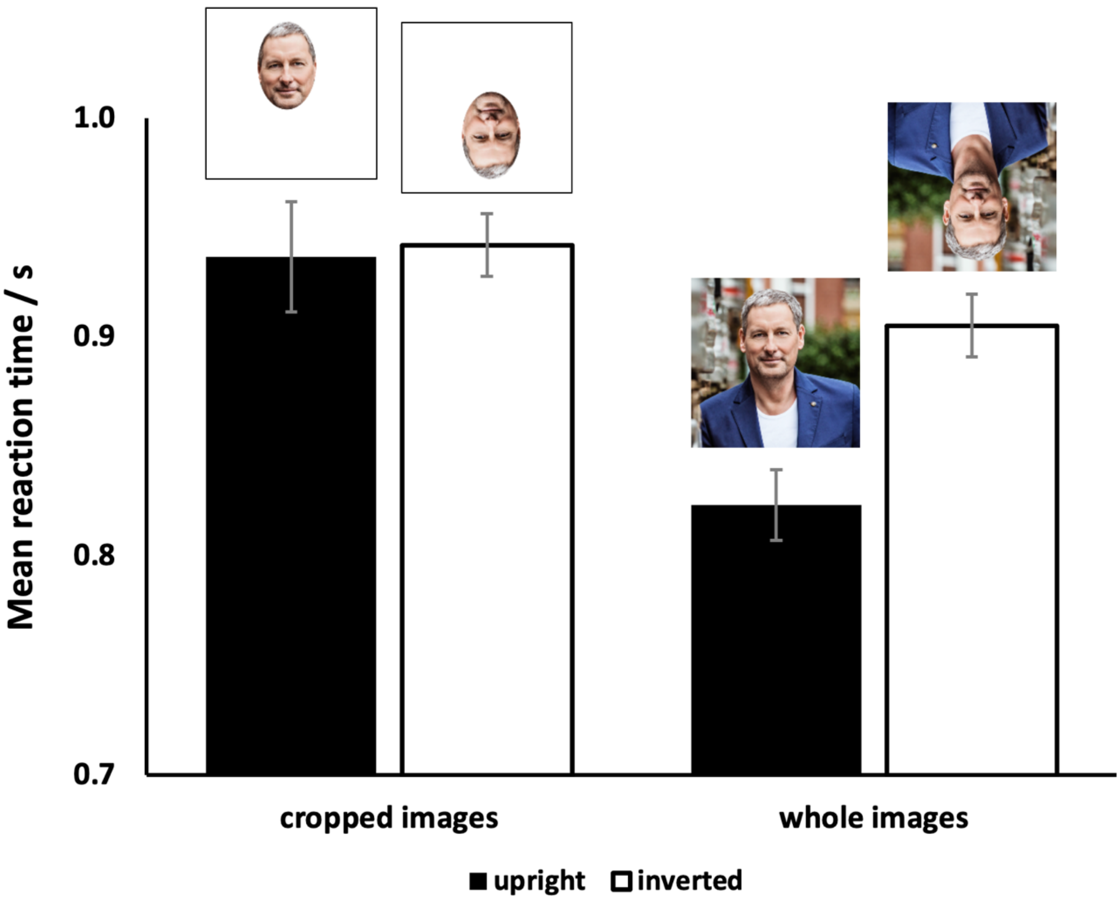



**Figure 3.5.10.** Mean reaction times for novel unfamiliar identities shown upright or inverted, and as cropped ovals or whole images.

**Discussion**

Whole and cropped images of learned unfamiliar faces produced near-identical sized inversion effects, both in accuracy (15.0% and 14.9% respectively) and in RTs (0.09s and 0.10s respectively). The proportion of participants showing an inversion effect was also almost identical for cropped faces (72%) and whole images (68%). This is in direct contrast to the experimental hypothesis, which predicted that whole images would suffer much smaller inversion effects compared to cropped faces. This is an interesting finding as the results are unambiguous, however, they are in opposition to the equally clear results found in Experiments 1 and 2.

Why then are these experiments producing such different results? There are 2 main differences between the experiments: the different tasks used and the familiarity of the faces. Although famous faces were the focus of Experiment 1 and 2, unfamiliar faces were also included as items for participants to respond, “no I do not recognise this face”. The unfamiliar faces in Experiment 1 and 2 reveal the same pattern of results found here in Experiment 5, that both cropped and whole images produced near-identical inversion effects. Therefore, as unfamiliar faces produced the same pattern of results across different tasks, the reason for the differing results between Experiments 1 and 2, and Experiment 5 could be due to the familiarity of the faces. The interaction between image cropping and inversion might be modulated by familiarity, with cropped familiar faces producing larger inversion effects than whole images of familiar faces, and both cropped and whole images of unfamiliar faces producing equal sized inversion effects. This would be an interesting and novel finding, however, such a conclusion cannot be strongly made as the idea is drawn from different experiments with varying tasks and participants. Therefore, Experiment 6 aimed to directly test for a modulatory role of familiarity.

# Experiment 6: Old-new memory task with familiar and unfamiliar faces

## Introduction

The face inversion effect has been replicated many times (McKone & Yovel, 2009), however, harshly cropped images of faces are often used in research. Experiments 1 and 2 used familiar identities to test whether the face inversion effect was stimulus-bound, whether it was due to the abnormal cropping of images used in experiments. They found the face inversion effect was reduced when using whole naturalistic images (familiar faces and buildings produced the same size inversion effects). However, the results of Experiment 5 were completely contradictory, with identical decreases in accuracy and increases in RTs for whole images and cropped faces. There were 2 key differences between Experiments 1 and 2, and Experiment 5: the familiarity of the faces and the different tasks used. This experiment aims to directly test if familiarity modulates the interaction between orientation and image type in one task. As in Yin (1969) and Experiment 5 here, an old-new memory task is used again, this time including famous and unfamiliar identities. It is hypothesized that for familiar IDs the pattern of results found in Experiment 1 and 2 will be replicated, that familiar faces will suffer a large decrease in accuracy when cropped harshly, and this will be drastically reduced when using whole naturalistic images. For unfamiliar identities, it is hypothesized that the same pattern of results found in Experiment 5 will be replicated, that both whole and cropped faces will produce equal sized inversion effects.

**Methods**

***Participants***

90 right-handed participants took part in this experiment (56 females, 33 males, 1 participant identified as “other”, mean age = 27.86, SD = 4.66, 2 participants did not provide complete dates of birth and are not included in the mean age or standard deviation). 23 additional participants were excluded from the analysis; 14 participants had missing cases, 5 participants got the buttons the wrong way round, 2 participants failed the attention check, and 2 participants experienced data recording issues. Participants were recruited from Prolific and were paid £5 to compensate their time. All participants self-reported as being right-handed, currently residing in the United Kingdom, had lived in the UK for at least 10 years, and had normal or corrected-to-normal vision. Informed consent was gained, and the experiment was approved by the University of York Psychology Ethics Board.

***Stimuli***

In this experiment, 96 images of unfamiliar IDs and 96 images of famous IDs were used. The 96 unfamiliar identities from Experiment 5 were used again in this experiment. The same square whole images from Experiment 5 were used, and of the oval cropped images, 74 of them were cropped more harshly to include only the internal features (chin, hairline and external cheek contour removed where needed). The same 48 famous identities from Experiments 2, 3 and 4 were used, however, new images for 2 of the identities were chosen to bear better resemblance, and 2 identities were swapped for different more readily recognisable identities. 48 more famous identities were collected, totalling 96 famous identities.

Each image was edited to create 4 versions: cropped upright, cropped inverted, full image upright, full image inverted. The whole images were cropped into squares, as in Experiment 5, and the cropped images were cropped into ovals including only the internal features, as shown in Figure 3.6.1. The images were fully counterbalanced across participants, so each participant only saw each identity once in only 1 of the conditions. (See Appendix G for further examples).

**Figure 3.6.1.** A sample of the stimuli used in Experiment 6. The top row shows familiar identities, and the bottom row shows unfamiliar identities. The left column

demonstrates harsh cropping around the internal features, and the right column the whole image.

### *Design and Procedure*

This experiment had the same design as Experiment 5 but with the addition of famous faces. Half of the unfamiliar and famous identities were shown during a learning task and half were used as novel distractors at test. In the learning phase, the images were all presented upright, and were either cropped ovals or whole images. At test, the items from the learning task and the novel distractors were presented either upright or inverted.

As this experiment had the same design as Experiment 5 with the addition of famous IDs, this experiment had 4 independent variables, each with 2 levels, producing 16 conditions. The familiarity level was either familiar or unfamiliar, the image type was either a cropped face or whole image of an identity, the orientation at test was either upright or inverted, and the faces were either learned or distractors. Participants completed 12 trials in each condition, 192 trials in total. Accuracy and reaction time were recorded.

The learning phase had the same design and procedure as in Experiment 5. As an attention check, participants again indicated via a button press (“F” or “M”) if faces were female or male (95.8% accurate) and were instructed to remember the identities. The order of trials was randomised for each participant. The learning phase lasted approximately 4 minutes. Participants then completed the classic Stroop task. This was identical to the Stroop task used in Experiment 5, with congruent and incongruent trials, presented either upright or inverted. This task lasted approximately 10 minutes. The memory task had the same procedure as Experiment 5, whereby participants pressed 1 or 0 to indicate if they saw the photo in the learning task, as shown in Figure 3.6.2. Again, the order of trials was randomised for each participant and the response keys were counterbalanced across participants. On average this task lasted approximately 5 minutes.

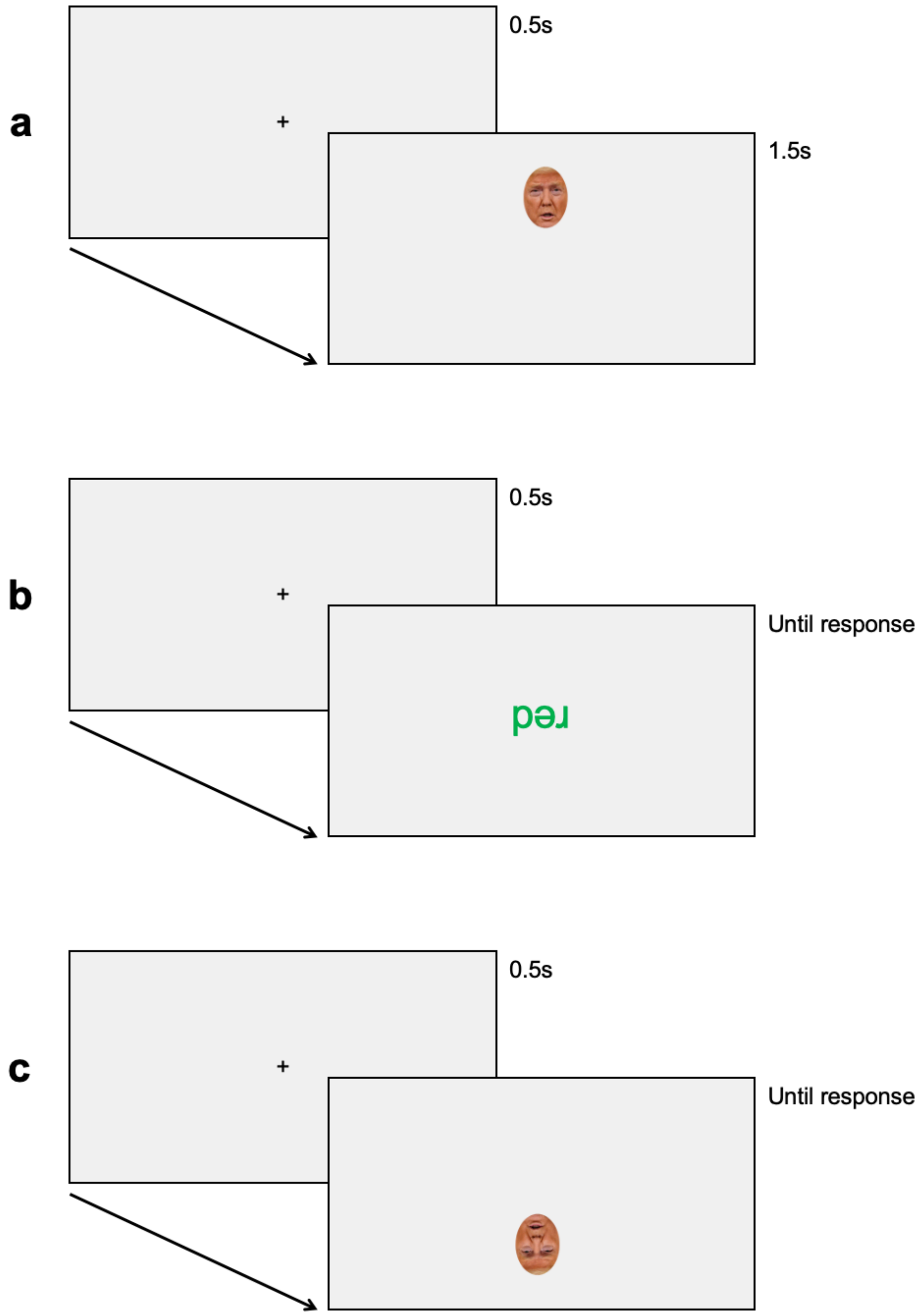


**Figure 3.6.2.** Example trials from Experiment 6. (a) In the learning phase, participants saw a fixation cross followed by an upright photograph. (b) In the Stroop task, participants indicated the colour a word was presented in while ignoring the meaning of the word. (c) In the test phase, participants saw a fixation cross followed by a photograph and had to indicate if they saw the person in the learning phase or not.

Following the memory task, participants completed a familiarity check task. Participants were presented with the upright full image photos of the famous identities and were asked to score how familiar they were with each identity by choosing one of 4 familiarity statements: 1 I could name the person, 2 I could say why the person is famous, but I cannot remember their name, 3 I recognise the person, but can't think why I know them, or 4 I do not recognise this person. Famous face trials in the memory test were discarded if participants rated the identity as a 3 or 4 in the familiarity check, which was taken as not recognising the famous face. This resulted in 11.6% of trials being eliminated on average for each participant (7 participants had more than 30% of trials binned, 9 participants had more than 20% of trials binned, 74 participants had below 20% of trials binned). Overall, this experiment lasted approximately 32 minutes.

## Results

*Stroop Task: Accuracy*

Figure 3.6.3 shows the mean accuracy. A 2x2 repeated measures ANOVA (congruent / incongruent; upright / inverted) was performed on the Stroop accuracy data. A significant main effect of congruency was found ($F(1,89) = 32.47$, $p < .001$, $\eta_p^2$

= 0.27), and there was no significant main effect of orientation (F(1,89) = 0.24, $p$ = .623, $\eta_p^2$ = 0.00). A significant interaction was found (F(1,89) = 13.45, $p < .001$, $\eta_p^2$ = 0.13). Simple main effects tests revealed that the interaction reflected the fact that inversion of congruent trials decreased accuracy by 1.0% ($p$ = .014), however, inversion of incongruent trials increased accuracy by 1.3% ($p$ = .003).

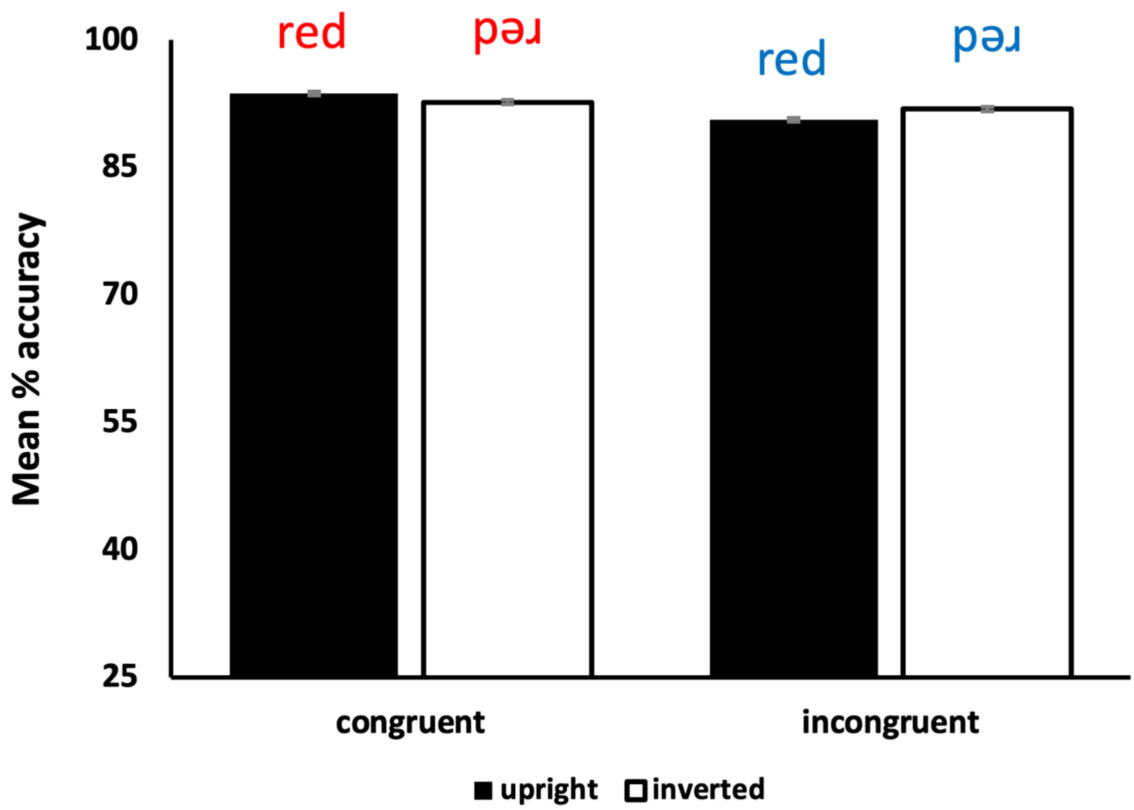



**Figure 3.6.3.** Mean percentage accuracy for upright and inverted, congruent and incongruent trials in the Stroop task.

*Stroop Task: RTs*

Figure 3.6.4 shows the mean reaction times. A 2x2 repeated measures ANOVA (congruent / incongruent; upright / inverted) was performed on the Stroop RT data. A significant main effect of congruency was found ($F(1,89) = 134.61$, $p < .001$, $\eta_p^2 = 0.60$), and there was no significant main effect of orientation ($F(1, 89) = 1.01$, $p = .319$, $\eta_p^2 = 0.01$). A significant interaction was found ($F(1, 89) = 40.95$, $p < .001$, $\eta_p^2 = 0.32$). Simple main effects tests revealed that the interaction reflected the fact that inversion of congruent trials increased RTs by 0.04s ($p < .001$) and inversion of incongruent trials decreased RTs by 0.03s ($p < .001$).

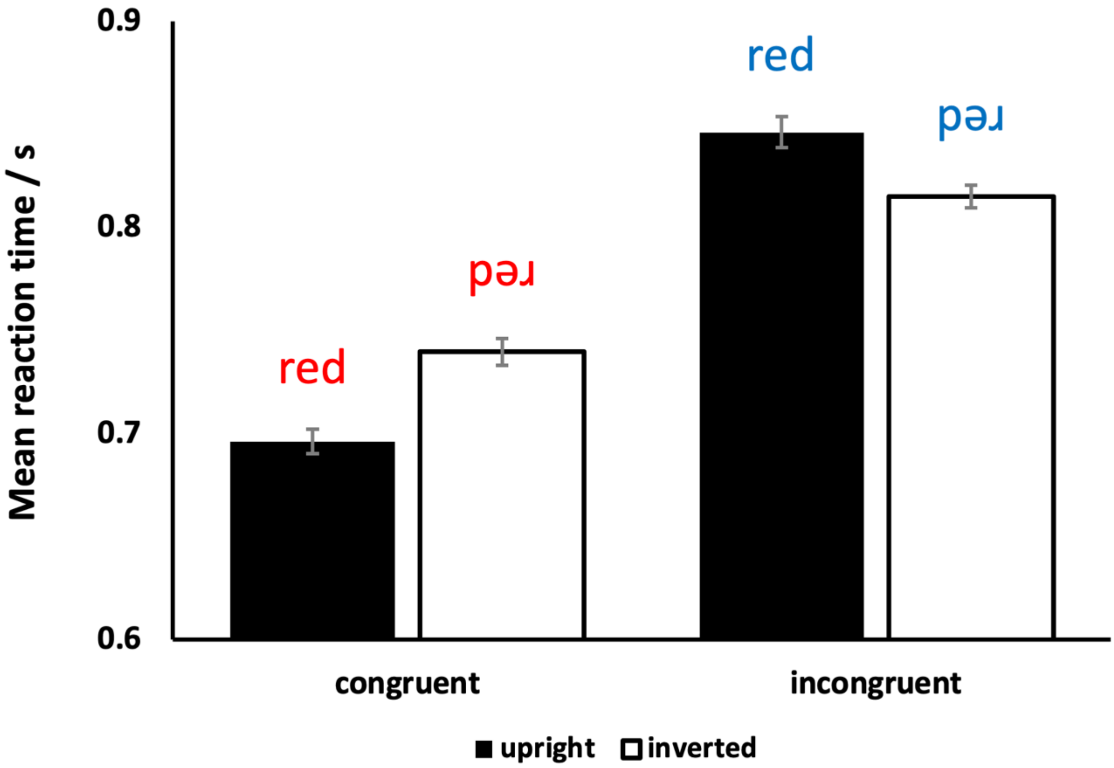



**Figure 3.6.4.** Mean reaction times for upright and inverted, congruent and incongruent trials in the Stroop task.

*Learned Famous and Unfamiliar Identities: Accuracy*

A 2x2x2 repeated measures ANOVA (famous / unfamiliar; whole image / oval cropped faces; upright / inverted) was performed on the accuracy data for the learned famous and unfamiliar identities. Significant main effects of familiarity ($F(1,89) = 214.89$, $p < .001$, $\eta_p^2 = 0.71$), image type ($F(1,89) = 18.74$, $p < .001$, $\eta_p^2 = 0.17$), and orientation ($F(1,89) = 72.24$, $p < .001$, $\eta_p^2 = 0.45$) were found. None of the two-way interactions were significant (familiarity * image type ($F(1,89) = 0.47$, $p = .494$, $\eta_p^2 = 0.01$), familiarity * orientation ($F(1,89) = 0.88$, $p = .351$, $\eta_p^2 = 0.01$), image type * orientation ($F(1,89) = 3.18$, $p = .078$, $\eta_p^2 = 0.03$)), however, the three-way interaction, familiarity * image type * orientation, was significant ($F(1,89) = 33.39$, $p < .001$, $\eta_p^2 = 0.27$).

*Learned Famous Identities: Accuracy*

Figure 3.6.5 shows the mean accuracy for learned famous identities. A 2x2 repeated measures ANOVA (whole image / oval cropped faces; upright / inverted) was performed on the learned familiar accuracy data. Significant main effects of image type ($F(1,89) = 17.87$, $p < .001$, $\eta_p^2 = 0.17$), and orientation ($F(1,89) = 53.65$, $p < .001$, $\eta_p^2 = 0.38$) were found, and a significant interaction was found ($F(1,89) = 28.55$, $p < .001$, $\eta_p^2 = 0.24$). Simple main effects tests revealed that the interaction reflected the fact that whole image inversion led to a 5.3% ($p = .004$) decrease in accuracy, whereas cropped face inversion led to a 20.6% ($p < .001$) decrease in accuracy.

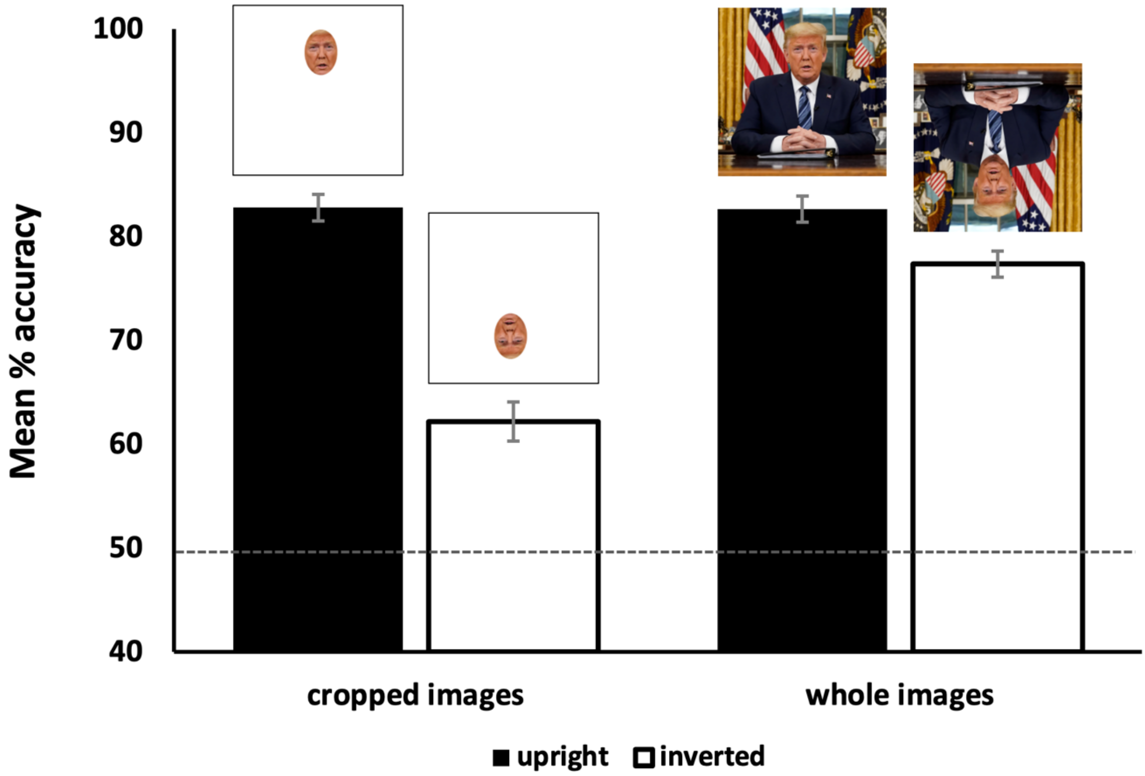



**Figure 3.6.5.** Mean percentage accuracy for learned familiar identities shown upright or inverted, and as cropped or whole images. The dashed line represents chance level.

*Learned Famous Identities: Inter-subject variability*

For the oval cropped faces, 81% of participants showed an inversion effect (their accuracy was lower in the inverted condition compared to the upright condition), 4% showed no effect (their accuracy was the same in the upright and inverted conditions), and 14% showed the opposite effect (their accuracy was higher in the inverted condition compared to upright), as shown in Figure 3.6.6. For the whole images, 59% of participants showed an inversion effect, 13% showed no effect, and 28% showed the opposite effect.

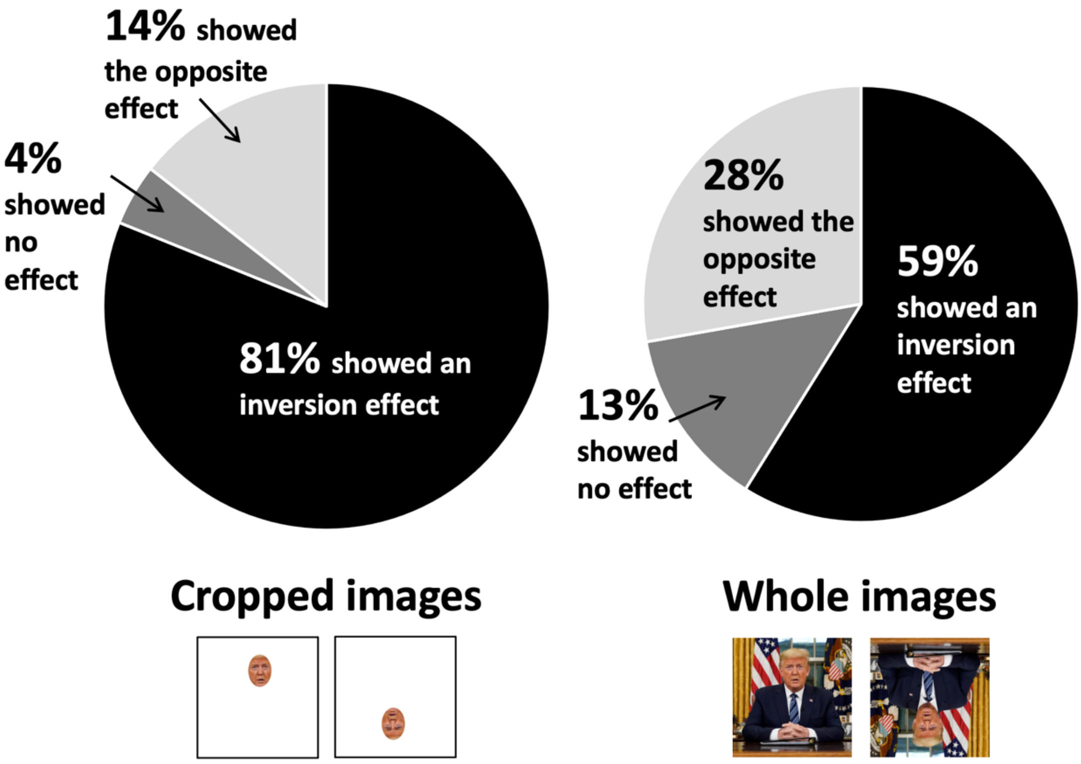



**Figure 3.6.6.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for learned famous identities shown as cropped or whole images.

*Learned Unfamiliar Identities: Accuracy*

Figure 3.6.7 shows the mean accuracy for learned unfamiliar identities. A 2x2 repeated measures ANOVA (whole image / oval cropped faces; upright / inverted) was performed on the learned unfamiliar accuracy data. Significant main effects of image type ($F(1,89) = 10.02$, $p = .002$, $\eta_p^2 = 0.10$), and orientation ($F(1,89) = 39.15$, $p < .001$, $\eta_p^2 = 0.31$) were found, and a significant interaction was found ($F(1,89) = 7.91$, $p = .006$, $\eta_p^2 = 0.08$). Simple main effects tests revealed that the interaction reflected the fact that whole image inversion led to a 15.0% ($p < .001$) decrease in accuracy, whereas cropped face inversion led to a 6.9% ($p = .003$) decrease in accuracy.

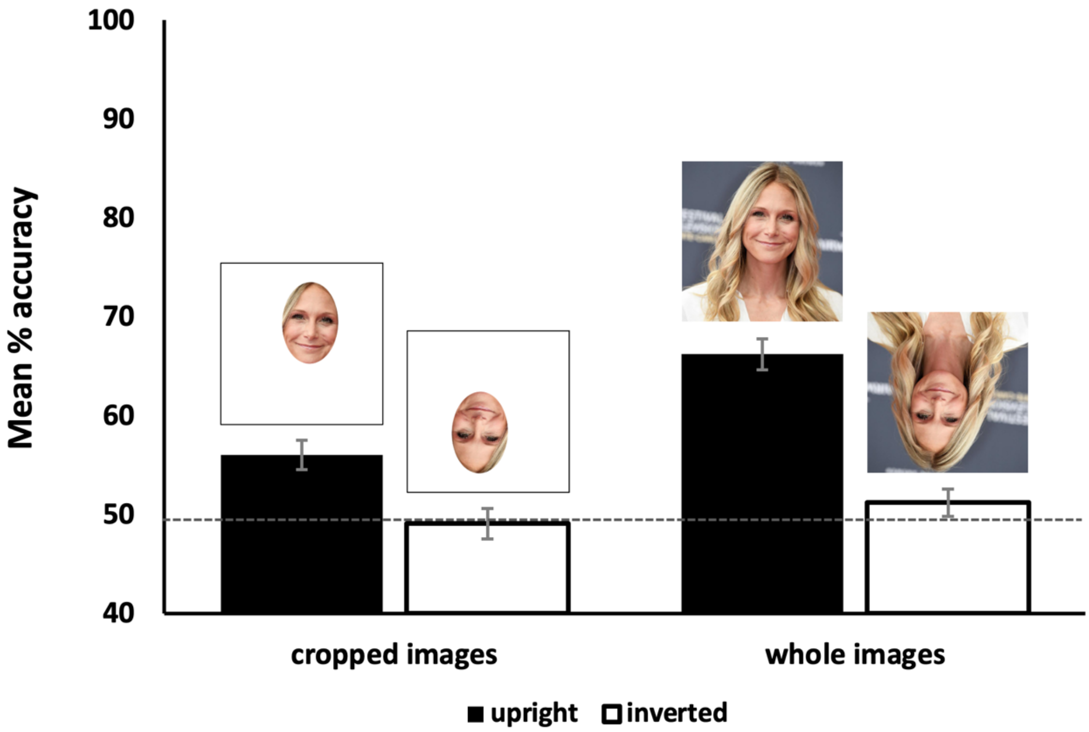



**Figure 3.6.7.** Mean percentage accuracy for learned unfamiliar identities shown upright or inverted, and as cropped or whole images.

*Learned Unfamiliar Identities: Inter-subject variability*

For the oval cropped faces, 56% of participants showed an inversion effect, 11% showed no effect, and 33% showed the opposite effect, as shown in Figure 3.6.8. For the whole images, 68% of participants showed an inversion effect, 13% showed no effect, and 19% showed the opposite effect.

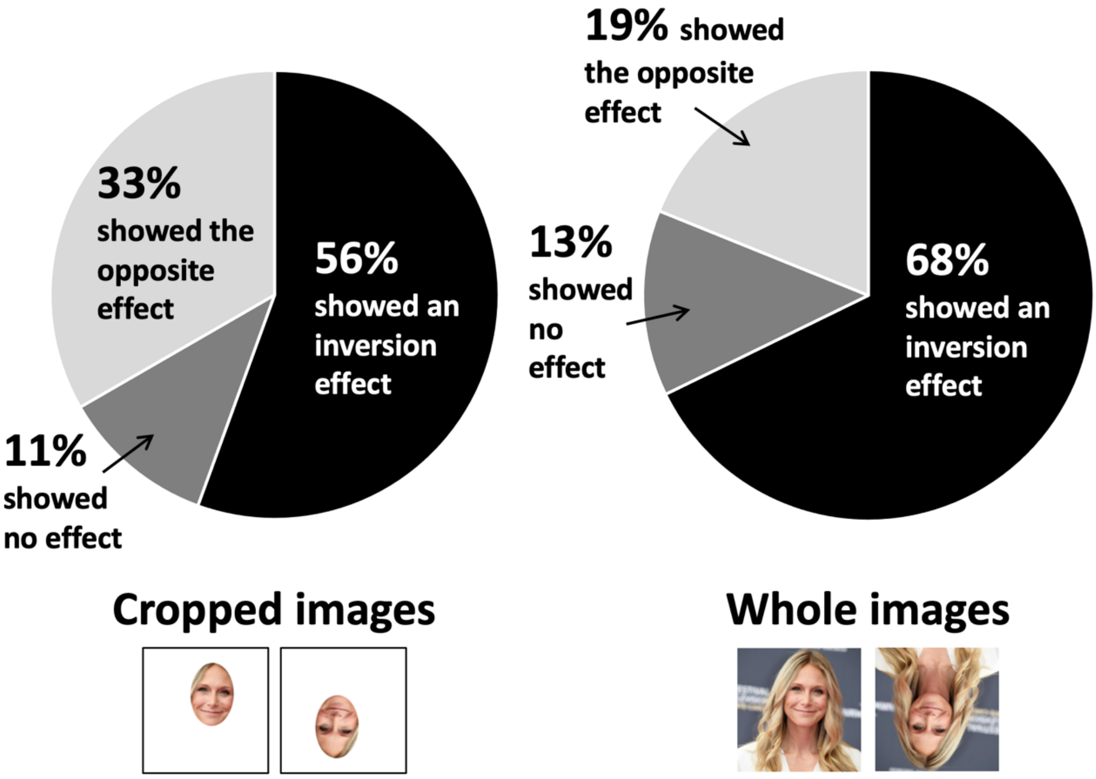



**Figure 3.6.8.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for learned unfamiliar identities shown as cropped or whole images.

*Learned Famous and Unfamiliar Identities: RTs*

A 2x2x2 repeated measures ANOVA (famous / unfamiliar; whole image / oval cropped faces; upright / inverted) was performed on the RT data for the learned famous and unfamiliar identities. Significant main effects of familiarity ($F(1,89) = 40.07$, $p < .001$, $\eta_p^2 = 0.31$), image type ($F(1,89) = 14.93$, $p < .001$, $\eta_p^2 = 0.14$), and orientation ($F(1,89) = 70.01$, $p < .001$, $\eta_p^2 = 0.44$) were found. None of the two-way interactions were significant (familiarity * image type ($F(1,89) = 0.01$, $p = .938$, $\eta_p^2 = 0.00$), familiarity * orientation ($F(1,89) = 2.40$, $p = .125$, $\eta_p^2 = 0.03$), image type * orientation

(F(1,89) = 0.88, $p$ = .088, $\eta_p^2 = 0.01$)), however, the three-way interaction, familiarity * image type * orientation, was significant (F(1,89) = 14.06, $p < .001$, $\eta_p^2 = 0.14$).

*Learned Famous Identities: RTs*

Figure 3.6.9 shows the shows the mean reaction times for learned famous identities. A 2x2 repeated measures ANOVA (whole image / oval cropped faces; upright / inverted) was performed on the learned famous RT data. Significant main effects of image type (F(1,89) = 10.47, $p$ = .002, $\eta_p^2 = 0.11$), and orientation (F(1,89) = 86.84, $p < .001$, $\eta_p^2 = 0.49$) were found; however, no significant interaction was found (F(1,89) = 2.84, $p$ = .095, $\eta_p^2 = 0.03$).

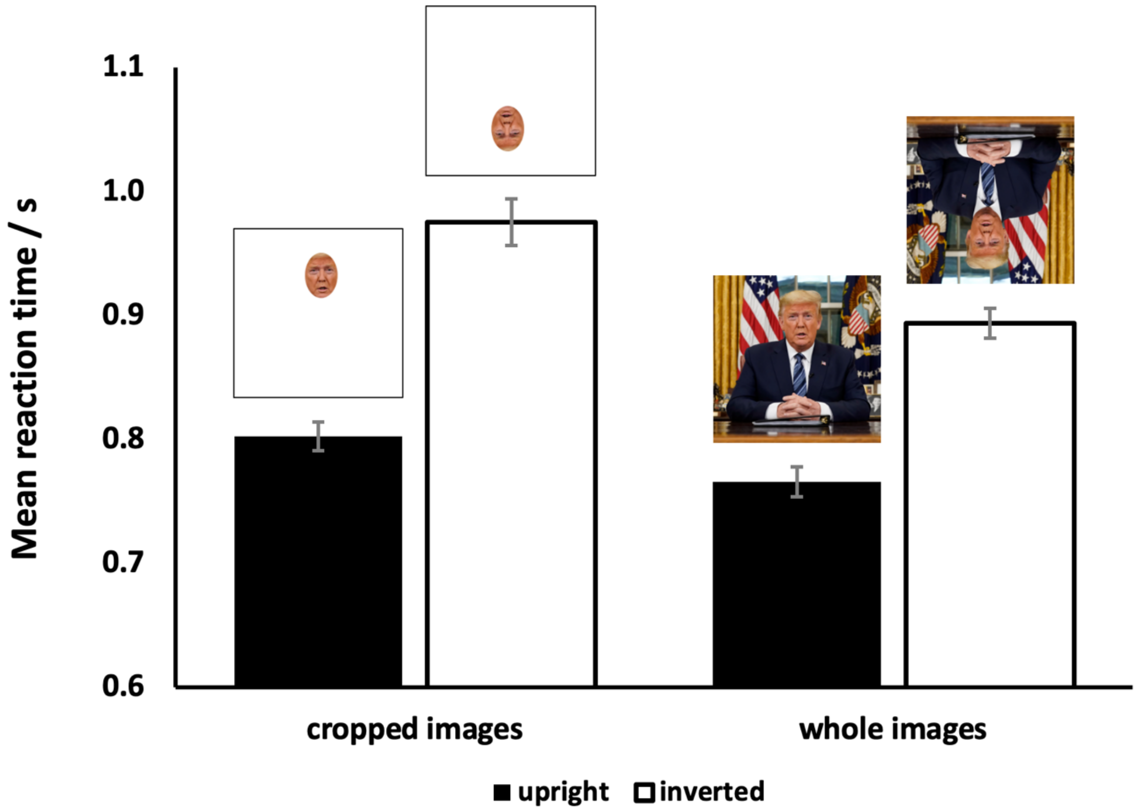



**Figure 3.6.9.** Mean reaction times for learned familiar identities shown upright or inverted, and as cropped or whole images.

*Learned Unfamiliar Identities: RTs*

Figure 3.6.10 shows the shows the mean reaction times for learned unfamiliar identities. A 2x2 repeated measures ANOVA (whole image / oval cropped faces; upright / inverted) was performed on the learned unfamiliar RT data. Significant main effects of image type ($F(1,89) = 9.10$, $p = .003$, $\eta_p^2 = 0.09$), and orientation ($F(1,89) = 31.15$, $p < .001$, $\eta_p^2 = 0.26$) were found, and a significant interaction was found ($F(1,89) = 9.08$, $p = .003$, $\eta_p^2 = 0.09$). Simple main effects tests revealed that the interaction reflected the fact that whole image inversion led to a 0.16s ($p < .001$) increase in reaction times, whereas cropped face inversion led to a 0.08s ($p = .008$) increase in reaction times.

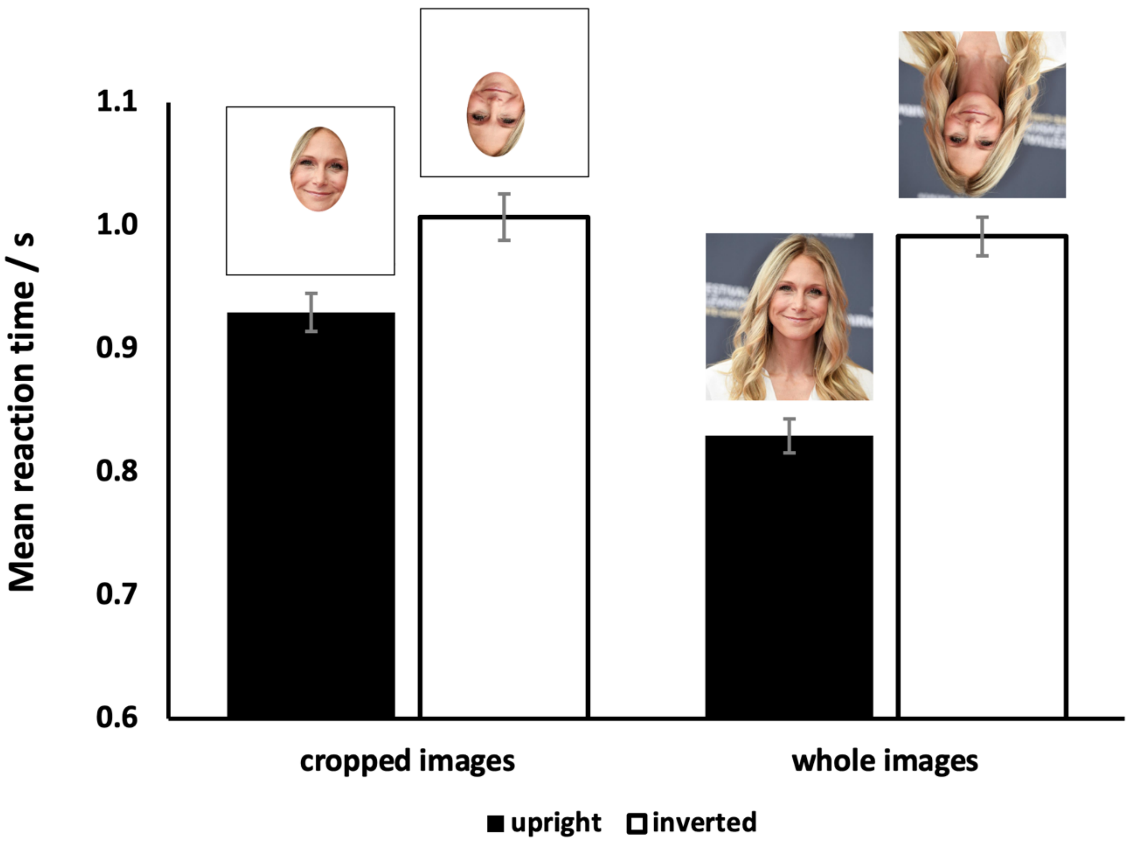



**Figure 3.6.10.** Mean reaction times for learned unfamiliar identities shown upright or inverted, and as cropped or whole images.

*Novel Famous and Unfamiliar Identities: Accuracy*

A 2x2x2 repeated measures ANOVA (famous / unfamiliar; whole image / oval cropped face; upright / inverted) was performed on the accuracy data for the novel famous and unfamiliar identities. Significant main effects of familiarity ($F(1,89) = 12.30$, $p = .001$, $\eta_p^2 = 0.12$), image type ($F(1,89) = 144.03$, $p < .001$, $\eta_p^2 = 0.62$), and orientation ($F(1,89) = 21.03$, $p < .001$, $\eta_p^2 = 0.19$) were found. There were significant two-way interactions between familiarity and image type ($F(1,89) = 9.06$, $p = .003$, $\eta_p^2 = 0.09$), and between familiarity and orientation ($F(1,89) = 10.99$, $p = .001$, $\eta_p^2 = 0.11$), however, the two-way interaction between image type and orientation was not significant ($F(1,89) = 0.63$, $p = .428$, $\eta_p^2 = 0.01$). The three-way interaction between familiarity, image type and orientation was significant ($F(1,89) = 5.97$, $p = .017$, $\eta_p^2 = 0.06$).

*Novel Famous Identities: Accuracy*

Figure 3.6.11 shows the mean accuracy for novel famous identities. A 2x2 repeated measures ANOVA (whole image / oval cropped face; upright / inverted) was performed on the novel famous accuracy data. A significant main effect of image type was found ($F(1,89) = 120.52$, $p < .001$, $\eta_p^2 = 0.58$). However, there was no significant main effect of orientation ($F(1,89) = 2.33$, $p = .130$, $\eta_p^2 = 0.03$) and no significant interaction was found ($F(1,89) = 3.84$, $p = .053$, $\eta_p^2 = 0.04$).


**Figure 3.6.11.** Mean percentage accuracy for novel familiar identities shown upright or inverted, and as cropped or whole images.

*Novel Famous Identities: Inter-subject variability*

For the oval cropped faces, 49% of participants showed an inversion effect, 7% showed no effect, and 44% showed the opposite effect, as shown in Figure 3.6.12. For the whole images, 48% of participants showed an inversion effect, 24% showed no effect, and 28% showed the opposite effect.

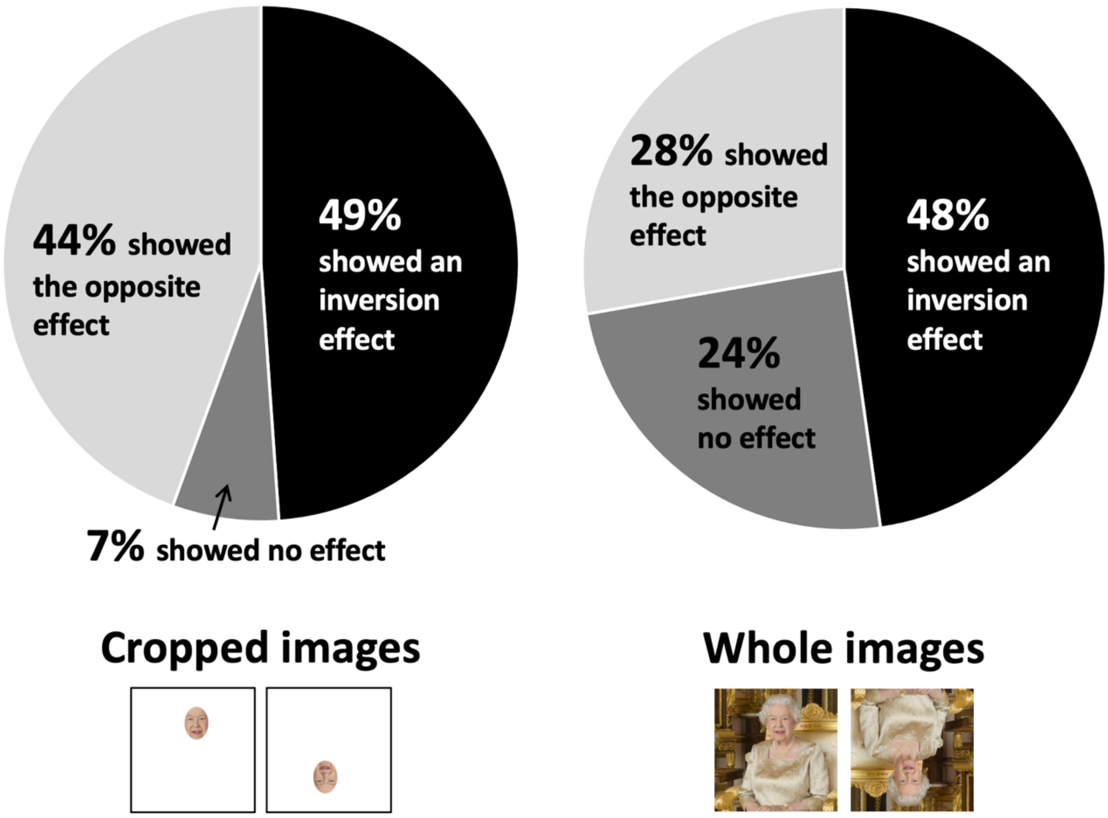



**Figure 3.6.12.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for novel famous identities shown as cropped ovals or whole images.

*Novel Unfamiliar Identities: Accuracy*

Figure 3.6.13 shows the mean accuracy for novel unfamiliar identities. A 2x2 repeated measures ANOVA (whole image / oval cropped face; upright / inverted) was performed on the novel unfamiliar accuracy data. Significant main effects of image type ($F(1,89) = 52.76$, $p < .001$, $\eta_p^2 = 0.37$) and orientation ($F(1,89) = 32.34$, $p < .001$, $\eta_p^2 = 0.28$) were found. However, there was no significant interaction ($F(1,89) = 0.94$, $p = .334$, $\eta_p^2 = 0.01$).

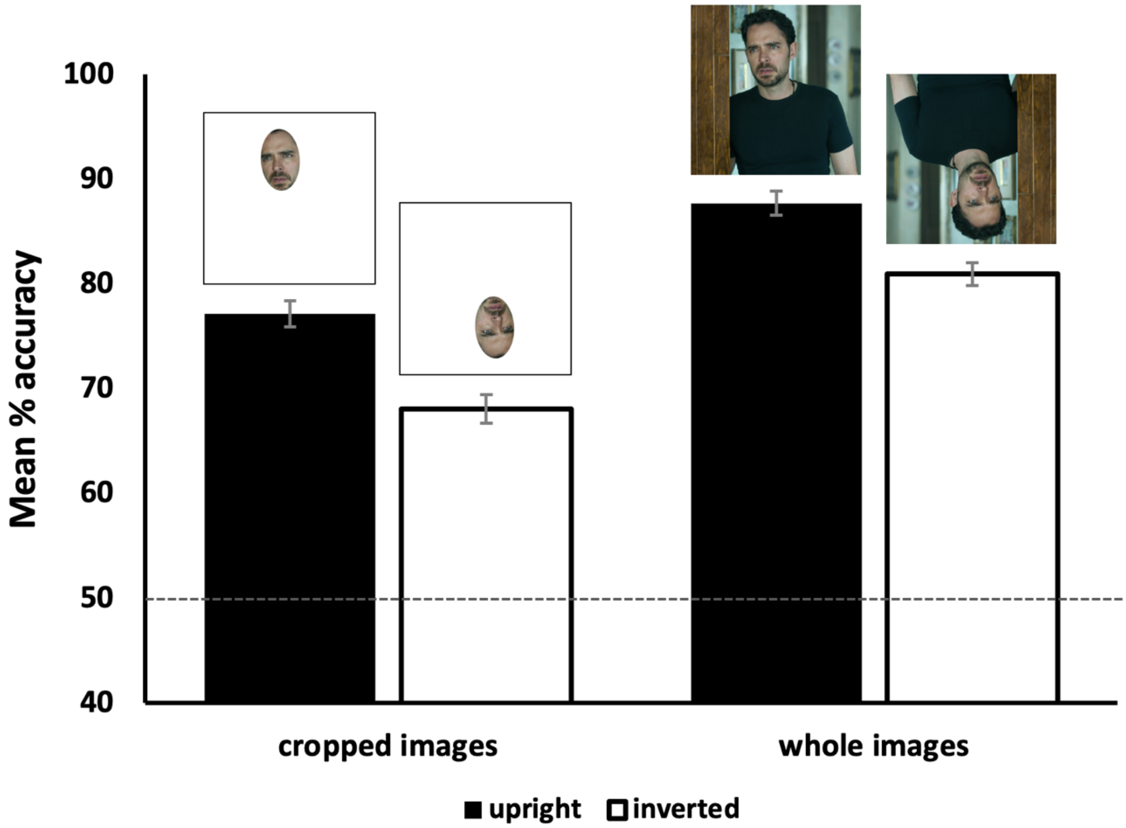



**Figure 3.6.13.** Mean percentage accuracy for novel unfamiliar identities shown upright or inverted, and as cropped or whole images.

*Novel Unfamiliar Identities: Inter-subject variability*

For the oval cropped faces, 54% of participants showed an inversion effect, 23% showed no effect, and 22% showed the opposite effect, as shown in Figure 3.6.14. For the whole images, 53% of participants showed an inversion effect, 27% showed no effect, and 20% showed the opposite effect.

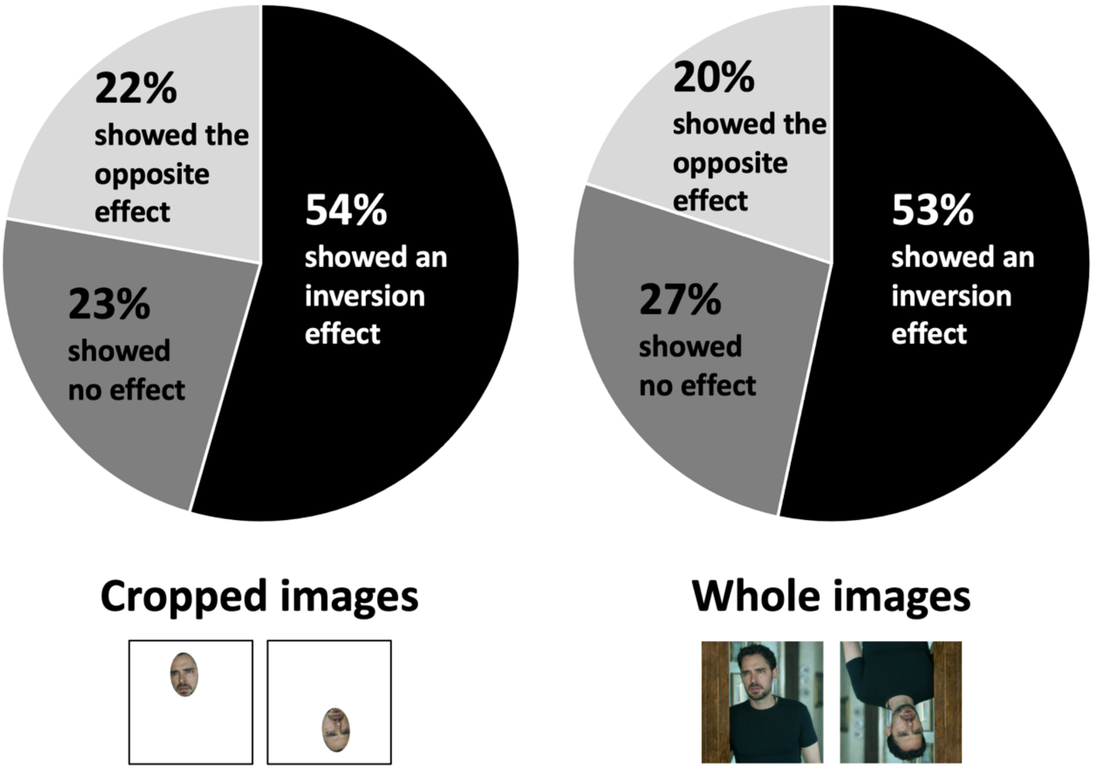



**Figure 3.6.14.** Pie charts displaying the percentage of participants who showed an inversion effect, participants who showed no effect, and participants who showed the opposite effect for novel unfamiliar identities shown as cropped or whole images.

*Novel Famous and Unfamiliar Identities: RTs*

A 2x2x2 repeated measures ANOVA (famous / unfamiliar; whole image / oval cropped face; upright / inverted) was performed on the RT data for the novel famous and unfamiliar identities. There was no significant main effect of familiarity ($F(1,89) = 0.48$, $p = .492$, $\eta_p^2 = 0.01$), however, there were significant main effects of image type ($F(1,89) = 19.08$, $p < .001$, $\eta_p^2 = 0.18$), and orientation ($F(1,89) = 29.94$, $p < .001$, $\eta_p^2 = 0.25$). There were no significant two-way interactions between familiarity and image type ($F(1,89) = 0.52$, $p = .474$, $\eta_p^2 = 0.01$), or between familiarity and orientation

(F(1,89) = 0.85, $p$ = .360, $\eta_p^2$ = 0.01) however, the two-way interaction between image type and orientation was significant (F(1,89) = 36.65, $p$ < .001, $\eta_p^2$ = 0.29). The three-way interaction between familiarity, image type and orientation was not significant (F(1,89) = 0.10, $p$ = .759, $\eta_p^2$ = 0.00). In order to interpret the significant two-way interaction between image type and orientation (as shown in Figure 3.6.15), simple main effects tests were run, and revealed that the interaction reflected the fact that whole image inversion led to a 0.11s increase in reaction times ($p$ < .001), whereas cropped face inversion led to a 0.01s decrease in reaction times ($p$ = .679).

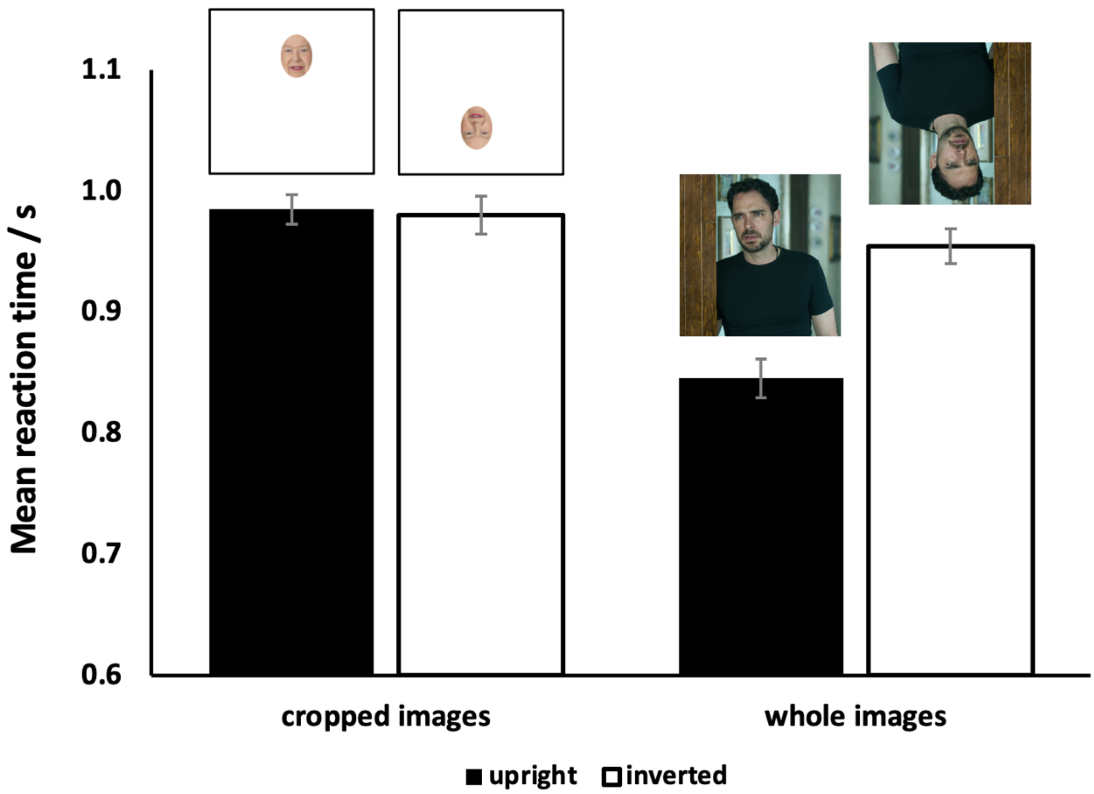



**Figure 3.6.15.** Mean reaction times for novel identities, collapsed across famous and unfamiliar identities, shown upright or inverted, and as cropped or whole images.

**Discussion**

In this experiment, a modulatory effect of familiarity was found on face inversion effects. For famous identities, using harshly cropped images of faces, large inversion effects (20.6%) were found; however, when naturalistic whole images of the IDs were inverted, much smaller decreases (5.3%) were found, replicating the pattern of results found in Experiment 1 and 2 in accordance with the hypothesis. Furthermore, more participants showed an inversion effect when looking at oval cropped faces (81%) compared to whole images (59%). The opposite pattern was found for unfamiliar identities. When harshly cropped images of faces were inverted, a small decrease in accuracy (6.9%) was found, with inverted faces being recognised at chance level (49.1%). However, larger decreases in accuracy (15.0%) were found when naturalistic whole images were inverted, again with inverted faces being recognised at chance level (51.2%). More participants got an inversion effect when looking at whole images (68%) compared to oval cropped faces (56%).

The pattern of data found for famous identities replicated the pattern found in Experiments 1 and 2. This further replication using a different task supports the hypothesis that for familiar identities, using whole images negates the severe impact of inversion. The pattern of data for unfamiliar identities however was not quite as expected. It was hypothesized that both harshly cropped and whole images would suffer equal size inversion effects, however, whole images produced larger inversion effects than cropped images. Initially this might seem like quite an unexpected result, however, closer inspection of the data reveals a less surprising finding. Firstly, the pattern of data does not provide evidence against a modulatory effect of familiarity, as the opposite, rather than the same, pattern was found as in the familiar face data. Furthermore, floor effects may have caused the interaction in the unfamiliar face data, and it is possible that if the data were above floor level, equal sized decreases in accuracy may have been found. This is plausible for a number of reasons. Firstly, the upright whole faces were identified as previously seen more accurately than the upright cropped faces – this makes sense as there was more information in the image to encode during learning and more information to cue memory at test. Whole images suffered a 15.0% decrease in accuracy hitting floor (51.2%), an identical decrease to that found in Experiment 5, so this could reflect a genuine effect not masked by floor effects. Since upright cropped faces were identified as previously seen less accurately than whole images, 56.0% compared to 66.2%, there was less ‘room’ for inverted cropped faces to decrease in accuracy before hitting chance level, thus resulting in only a 6.9% decrease rather than a similar sized decrease to the whole images. Near identical sized decreases in accuracy for cropped and whole images of unfamiliar identities were seen in Experiments 1, 2 and 5, and therefore these results help to clarify the results here masked by floor effects.

# General Discussion

The results of this chapter reveal new findings about the face inversion effect. In Experiment 1, when using familiar faces cropped around the external features, the classic face inversion effect was found in the accuracy data: faces suffered a much larger decrease in accuracy (19.2%) compared to buildings (6.4%). However, this effect was reduced when using whole naturalistic images: familiar identities and buildings suffered similar sized inversion effects (6.3% and 7.3% respectively). Although image cropping did not have an effect in the RT data, reaction times to faces were slowed down more for inverted faces compared to buildings. Experiment 2 sought to replicate and extend this novel finding. In Experiment 2, the same pattern was found again, that cropped heads suffered much larger decreases in accuracy (27.6%) compared to whole images (15.1%). However, Experiment 2 also demonstrated that it was not the presence of a meaningful background which cued identity and led to the reduction in inversion effects, as images cropped around the body produced near-identical decreases in accuracy (16.3%) compared to the whole images (15.1%), and identical increases in RTs (0.16s for both whole bodies and whole images). Experiment 3 sought to replicate the finding of image cropping using a different task (a name-face verification task), however, this task proved to be too easy and produced ceiling effects. However, the RT data showed no effect of image cropping. Similarly in Experiment 4, an attempt was made to replicate the finding using an old-new memory task, however, due to experimental design flaws and the ease of the task, any pattern was again masked by ceiling effects, and similar increases in RTs were observed for both whole and cropped familiar faces.

Experiment 5 was a further attempt at improvement on replicating the finding using an old-new memory task. In an effort to increase the task difficulty, unfamiliar identities were used. This time, both cropped faces and whole images of unfamiliar IDs produced equal sized inversion effects (14.9% and 15.0% decreases in accuracy, 0.10s and 0.09s increases in RT respectively). This was unexpected following the results of Experiment 1 and 2 which used familiar identities. In order to test whether familiarity was modulating the effect of image type on orientation, familiar and unfamiliar identities were used in an old-new memory task in Experiment 6. The pattern found in the accuracy data of Experiments 1 and 2 using familiar IDs was reproduced here: cropped faces suffered a larger decrease in accuracy (20.6%) compared to whole images (5.3%). While Experiment 2 did find a modulating effect of image type in the RT data, the present experiment, like Experiment 1, showed no significant differences in the RT data. Unfamiliar identities produced similar results to that of Experiment 5, however, the pattern was partially masked by floor effects. Whole images of identities, as in Experiment 5, produced a 15% decrease in accuracy, however, cropped faces produced a smaller (6.9%) decrease in accuracy. Similarly, whole images showed larger increases in RTs due to inversion compared to cropped images. This was likely due to the low accuracy which the upright cropped faces produced, therefore participants did not have much ‘room’ to decrease in accuracy before hitting chance level. Overall, the accuracy data of these experiments demonstrate that inversion effects are modulated by the image cropping and the familiarity of the identity. For familiar IDs, inverting cropped faces produced larger inversion effects than naturalistic whole images. However, for unfamiliar identities, the cropping of the image had no effect, and similar sized inversion effects were found for cropped faces and whole images of identities.

Some of these results fit with the existing FIE literature, and some are controversial. Looking first at the cropped unfamiliar faces as these are the stimuli most often used in FIE experiments (Bombari et al., 2009; Civile et al., 2016; Rakover et al., 2022; Sekuler et al., 2004), the results are consistent with the literature. Experiment 1 replicated the classic FIE, whereby cropped unfamiliar faces suffered a larger decrease in accuracy when inverted (9.1%) compared to cropped unfamiliar buildings (1.7%). This result is consistent with the larger inversion effects for faces compared to non-face stimuli found in Yin's (1969) seminal paper, and with Diamond and Carey (1986) who also found a larger decrease for faces (19%) compared to landscapes (9%). Furthermore, similar sized inversion effects were obtained for unfamiliar faces whether harshly cropped faces or whole images were used. This supports the validity of the many studies which used cropped images of unfamiliar identities, as similar results may have been obtained if whole naturalistic images were used. Although Yin (1969) does not report reaction time data, Experiment 4 of Diamond and Carey (1986) does. They found that faces were overall responded to faster than non-face stimuli (landscapes and houses), and they also found that inversion slowed down reaction times equally for face and non-face stimuli, consistent with the results found in Experiment 1 here.

The results from the cropped familiar faces also support the literature. Marzi and Vigianno (2007) also found larger decreases in accuracy when cropped familiar faces were inverted, compared to smaller decreases for unfamiliar faces in a recognition task. They also found larger increases in RTs when inverting cropped familiar faces compared to inverting cropped unfamiliar faces, consistent with the pattern of results found in Experiment 1, 2 and 6 here. Similarly, Wang et al. (2023) found large inversion effects for personally familiar faces but no inversion effect for unfamiliar faces. The results of the present chapter support the findings of these studies, that when using cropped stimuli, familiar faces suffer larger inversion costs than unfamiliar faces in both accuracy and reaction times. However, the results from the whole images of familiar faces produced novel findings.

To the best of my knowledge, this chapter contains the first attempts at comparing cropped and whole images of upright and inverted familiar faces. The large decrease in accuracy found when using cropped images of inverted familiar faces can be minimised by using whole images, and, as shown in Experiment 1, when using naturalistic whole images of familiar faces and buildings, they produced equal sized inversion effects. However, the reaction time data reveal a less clear picture. In Experiment 1, familiar faces were recognised faster overall than familiar buildings and inversion increased RTs for faces more than buildings. However, image cropping did not affect RTs for both familiar faces and buildings, even though image cropping did affect accuracy. The accuracy data represents a novel finding which is at odds with the FIE literature which finds larger inversion effects for faces compared to non-face stimuli.

However, these results do fit with recent research which finds no face inversion effect. Gerlach (2023) found similar sized inversion effects for both faces and objects, by using a within-category discrimination task for faces and an object decision task for objects. They argued that FIE experiments typically require participants to make within-category discriminations for both faces and objects, however, we are only experienced at performing that level of discrimination with faces. They argued that if the experimental task matches the type of processing usually performed on a stimulus type, inversion effects would also be found for non-face objects. The present results might initially seem at odds with this paper as, in Experiment 1, a within-category discrimination task was used for both people and buildings (a recognition task). However, for whole images, equal size inversion effects were found for familiar identities and buildings. This may be because for familiar buildings (in this experiment, but perhaps objects in general) individual level recognition *is* required in day to day life. For example, when you enter your place of work, you do not simply recognise the visual stimulus as ‘a building’ but recognise the specific building, ‘my workplace’ (Wiese et al., 2023). Therefore the present results fit with Gerlach (2023), that the face inversion effect is reduced, that faces and non-face objects suffer similar inversion effects, when the experimental task reflects the real world processes typically performed on such stimuli.

Why then do I find results consistent with the literature when using unfamiliar identities regardless of image cropping, but find contradictory results with familiar identities when the images are uncropped? The answer may lie in the vast differences between unfamiliar and familiar faces. It has been suggested that unfamiliar faces are not faces, but instead are processed in the same way as non-face stimuli (Megreya & Burton, 2006). We are very poor at matching unfamiliar faces even without time limits, with studies typically finding 30% errors (Bruce et al., 1999). We are slower and less accurate at recognising unfamiliar faces (Clutterbuck & Johnston, 2005) and perceive multiple photographs of one unfamiliar identity to be photographs of many different people (Jenkins et al., 2011). Our expertise however lies with familiar face recognition. We are able to tell many variable photographs of a familiar face together and recognise them as one person (Jenkins et al., 2011), and we are very accurate and fast at recognising familiar faces in an effortless automatic process (Clutterbuck & Johnston, 2005).

These large differences between familiar and unfamiliar face processing are due to the underlying face representations. As discussed in Chapter 1, our face representations form as we observe idiosyncratic within-person face variation of an individual. Faces vary across many dimensions, e.g. viewpoint, emotional expression, speech, lighting etc. (Jenkins et al., 2011). The observed within-person variation is stored into a person-specific robust face representation (Bruce, 1994). Novel instances of familiar faces can then be recognised accurately because the face representation contains enough within-person variation to be able to generalise and the novel instance is accurately matched to the face representation (Jenkins & Burton, 2011). In stark contrast, purely unfamiliar identities (never seen before faces) do not have a face representation, as they have not been seen before so there is no visual information to store. Therefore, it is impossible for unfamiliar faces to be 'recognised'. In experiments when two photographs of unfamiliar faces are being matched, as they have no face representation they are not being recognised, therefore a simple image matching strategy is used instead, producing the typical poor performance associated with unfamiliar identities (Hancock et al., 2000). In memory experiments using unfamiliar faces, when the unfamiliar face is seen in an exposure phase, a face representation starts to be formed. The identity moves up the familiarity continuum from purely unfamiliar (never seen before) to highly unfamiliar (a small unstable face representation containing just 1 instance of a face) (Kovács, 2020). Therefore, unfamiliar faces can be 'recognised', as there is stored visual information about the face which an image can be matched to, however this is of an apparently qualitatively different nature to familiar face representations.

However, this theory of underlying face representations and the Face Recognition Units (FRUs) in Bruce and Young's (1986) model only considers faces. This is likely because research has focused on identification from the face, and has often been conducted on photographs of faces cropped around the internal / external facial features. However, the face makes up only approximately 3.5% of the surface area of a person (Liu et al., 2008). When stood in close proximity to someone, for example during a conversation, the visual information observed is not just their internal features, but also their whole head and upper body, and it is this information which forms the face representation. Faces never appear as cropped ovals apart from in experimental research. Therefore, I propose that a more useful construct may involve extending the idea of a visual *face* representation to incorporate the body, becoming a visual *person* representation. Or to use the language of Bruce and Young's (1986) model, a *face* recognition unit (FRU) becoming a *person* recognition unit (PRU) – not to be confused with person identity nodes which refers to the semantic information about the person.

This updated theory of person recognition (as opposed to face recognition) still places greater weight on the face over the body for recognition. This is because the majority of the time spent with people is in close proximity where the face and only the upper body (e.g. shoulders) will take up the majority of the visual field, for example when ordering a coffee or talking with work colleagues. When people are at a distance, e.g. crossing a street to start a conversation, the body occupies the majority of the visual field and visual acuity is worse, therefore in these less frequent situations, the body would be weighted more for recognition over the face (Hahn et al., 2016). In this sense, this update to our theory of face (person) representations and upright face recognition does not pose a large change to current theory nor its predictions, but an important adjustment to include and appropriately weight the body in face (or person) representations to better reflect our visual experience in the world rather than in the lab. Such an adjustment has small implications for upright face recognition, as our ability to generalise and recognise familiar faces is excellent, however, it might have larger implications for inverted faces.

Evidence exists which supports the idea of a visual person representation over a face representation. When discussing the neuroanatomy of the face selective regions, the FFA is often discussed. However, the body-selective fusiform body area (FBA) is situated immediately next to the FFA (Peelen & Downing, 2005) and in fact might overlap with the FFA (Schwarzlose et al., 2005). The FFA is often said to be face-selective, however, although its most preferred stimulus type is faces, it also responds quite strongly to bodies, and to bodies more than to tools, scenes and other stimuli (Peelen & Downing, 2005; Pitcher et al., 2019). Research shows that faces and bodies in their correct configuration (a head above a body) are integrated and processed as one item rather than two separate items (Bernstein et al., 2014). Furthermore, fMRI adaptation studies have shown that prolonged viewing of a headless body leads to an identity adaptation effect in the face (Ghuman et al., 2010), and further research demonstrates that distributed identity representations are shared by both faces and bodies in a person identity representation rather than a face identity representation (Foster et al., 2021). The growing body of evidence suggests that the face is not processed in isolation but is inextricably linked to the body.

The results of this chapter fit with such an explanation. In Experiment 1, looking first at the recognition accuracy of the upright familiar identities, both the cropped and whole images produced very high accuracy, potentially ceiling effects. For the whole images, the photograph contains the type of visual information stored in the visual person representation (a head attached to a body, no part of the body cropped out). Therefore, when participants saw images in this condition, they scanned their visual person representations, and as the person representation is very robust and the image is typical of the type of image in the representation, the representation could easily generalise to match the incoming image, and the participant correctly recognised the person. For the cropped head, although the image is not typical of the kind of visual information stored in the representation (the incoming image has the body missing), as greater weight is given to faces for recognition and the representation is robust enough to generalise and match the incoming image, participants still accurately recognised the person. This explanation also fits with the upright familiar face RT data, that both were also recognised very quickly, and whole images were matched slightly faster than cropped images.

For the inverted whole image, the photograph still resembles the content of the visual person representation (a whole body), however, the image is inverted and the vast majority of instances in the visual person representation are upright. It is therefore harder to generalise, and participants suffered a small percentage of errors and increase in RT. For the cropped inverted faces, a lot of the information is missing from the image in addition to the image being inverted, rendering the incoming image not at all like the visual content of the person representation. These two factors (image cropping and inversion) may have an interactive effect on face recognition accuracy– the percentage of errors caused by inverting a cropped image is greater than the sum of the effect of inverting an image plus the effect of cropping an image. An inverted cropped head is not at all similar to the visual input received and stored in the visual person representation and therefore the representation struggles to generalise to match the image, resulting in large errors and longer RTs typical of the kind of results found in FIE research. However, the increase to reaction times were equivalent when inverting cropped and whole images of familiar faces, which is harder to reconcile with such an explanation. However, as familiar face recognition is still a very robust process, and the experimental task being performed on the familiar IDs (recognition) is very typical of the processing done with familiar faces in day-to-day life, performance was still above chance.

The pattern of effects here are modulated by both image cropping and familiarity. One explanation for the pattern of results could be that the stimuli used are more realistic, and therefore reduce the inversion effect. However, this would not explain why familiarity also modulated the effect, why using more realistic stimuli does *not* reduce the inversion effect for unfamiliar faces. It is the combination of image cropping and familiarity both modulating the effect of inversion which may help to explain the results. For both familiar and unfamiliar faces, the whole uncropped images look more realistic and are therefore encoded more accurately into face space. For familiar identities, the whole images looking more realistic is the equivalent to the body aiding with identification. This is because, as described above, a whole image is encoded into face space closer to the other previously seen instances compared to a cropped image which would be encoded into face space further away. This then leads to more accurate recognition of the identity. For unfamiliar faces, however, there are no previous observed instances of the face and therefore the more realistic whole images do not aid with recognition.

The pattern of results found in this chapter are complimentary with Valentine's (1988) multidimensional space framework. His model suggests that dimensions can represent various aspects of faces, and when combined they create a multidimensional space where a face can be represented by a specific point. He suggests that an incoming image is encoded with a degree of error and, in the exemplar-based model, the Euclidian distance is calculated between the location of the incoming image and its nearest neighbour to aid recognition. The model states that inversion would increase the encoding error. For (typical) unfamiliar faces, the model predicts that the increased encoding error would produce a false-positive, as the encoded point would be closer to another face's representation and it would be misidentified as familiar. This fits with the pattern of data found here for unfamiliar faces in each experiment, that accuracy decreased for inverted unfamiliar faces, i.e. participants were more likely to incorrectly say they recognised the face. For familiar faces, Valentine's model predicts that inversion would again increase the encoding error, producing a point in face space further away than the location of the stored face representation, resulting in misses. This also fits with the data reported here, that accuracy decreased for inverted images of familiar faces across all experiments.

The proposed explanation of visual person representations is also similar to Valentine's (1991) model, with some differences. In Valentine's model, a face is represented in multidimensional space by 1 point, and a move away from that would represent a change in identity. However, in my explanation, as within-person variation of an individual is observed, multiple near-by points would be encoded and a region of multidimensional space would represent the person. Furthermore, in his model Valentine states that incoming faces are encoded into the space with a degree of error, and inverting the image increases this error, resulting in the encoded image being far away from the target location. However, in my explanation, the incoming image is not far away from the target location purely because of encoding error, but primarily because the incoming inverted image looks different from the previously encoded instances of the individual and is therefore plotted at the edge of the cluster of previously encoded instances, i.e. the person representation. In the present results, the inverted whole images of familiar faces look slightly dissimilar to previous instances and so are encoded slightly further away from the cluster, whereas cropped inverted images look highly dissimilar and are encoded much further away from the cluster. Both my explanation and Valentine's model suggest that the distance between the encoded stimulus and the nearest neighbour is calculated for recognition.

The question often asked of face inversion research is 'are faces special'? Much of the FIE literature revolves around trying to find the mechanism which is inhibited by inversion and therefore is thought to drive upright recognition. However, the results of Experiment 1 and the explanation of visual person representations do not provide strong evidence that faces are special. In the whole image condition, the condition which best represents real life experience and normal visual input, both familiar identities and familiar buildings suffered similar sized inversion effects (6.3% and 7.3% respectively) (However, whole images of familiar faces suffered larger increases in RTs when the images were inverted compared to whole images of buildings). The accuracy data is evidence which goes strongly against the FIE literature and suggests that there is *not* a factor specific to face recognition which causes large decreases in accuracy when faces are inverted (Yin, 1969), however the pattern in the RT data is less clear. For the cropped stimuli however, a larger decrease in accuracy and larger increase in RTs for faces compared to buildings was observed. However, I suggest that this does not provide evidence that faces are special, but provides evidence that faces are especially susceptible to poor stimuli design by researchers. In an earlier section I noted that the external features of faces are often cropped out in research, but an equivalent crop is not applied to control stimuli (for example, Ashworth et al. (2008)). However, even in Experiment 1 where I cropped around the external features of the face (including the hair and any visible external feature of the face) to match the external feature crop applied to buildings, this is still not a fair comparison. Such a crop leaves approximately 3.5% of the total surface area of a person remaining, however, none of the building is cropped out. The crop applied to buildings in Experiment 1 is actually more akin to the crop applied to the whole-body condition in Experiment 2. Much of the person is missing from the image in Experiment 1 but none of the building is. So therefore, it is not a fair comparison – however, what is? I think this question is the result of the different treatments given to face and non-face stimuli in research. People are made up of a head and body, and so the body can easily and intuitively be cropped out of an image. However, buildings do not have an equivalent to a face and body, and so there is no logical place to crop and remove the majority of the image. Therefore, researchers have removed the majority of the person from the image when studying faces, but left control stimuli intact, so the face stimuli do not resemble the familiar visual person representations, but the building stimuli do resemble the familiar visual building representations.

In Experiment 1, this led to large decreases for faces but not buildings. However, I hypothesize that if only 3.5% of the surface area of a building was used as a stimulus, whether presented upright or inverted, performance would be at chance level and RTs would be long, however, upright cropped faces were recognised with 92% accuracy, and decreased by approximately 20% when inverted to 73% in Experiment 1, and were recognised on average in 0.8s which increased by approximately 0.17s to 0.96s. If such results were found, this would suggest that familiar faces are special, but not in the way originally thought. As such severely cropped building stimuli would likely sit at floor, the stimulus type * orientation interaction could not be tested. Nevertheless, such comparatively high accuracy and fast RTs for familiar faces would demonstrate excellent familiar face recognition and its ability to withstand vast changes and image distortions (body removed and image inversion), something likely specific to faces.

For unfamiliar identities, again looking first at the upright image conditions, when participants viewed images in both the cropped and whole image conditions, their task was to indicate whether they recognised the person in the photograph or not. Therefore the cognitive task which participants carried out was in relation to familiar faces. In order to say whether they recognised an unfamiliar person or not, participants scanned their familiar visual person representations. With familiar people, we are excellent at telling them together because we have representations which can generalise and recognise a new instance very accurately. As the photo was of an unfamiliar person, none of the representations generalised to match the photo, regardless of whether the photograph was a cropped head or a full image, and therefore participants could very accurately reject the image as unfamiliar and begin forming a new representation of the identity. Unfamiliar face perception is often thought to be very poor, however, performance depends on the task and therefore the underlying cognitive processes. For example, matching tasks with unfamiliar faces do not rely on recognition via robust visual person representations but rather simple image matching strategies, and therefore performance is often poor (Hancock et al., 2000). However in Experiment 1, the task and therefore the underlying cognitive process was in relation to familiar faces and their robust representations, and therefore performance was high and RTs were quick. When the images were inverted, this led to more errors and longer RTs. It is known that everything is harder to recognise upside down, however, here the inverted unfamiliar faces are ‘easier to recognise’ upside-down, or rather they are incorrectly ‘recognised’ as familiar more often. The same pattern was also found with inverted unfamiliar buildings. This suggests that inversion makes the visual system more error prone, rather than specifically inhibiting recognition, and therefore caused participants to ‘recognise’ inverted unfamiliar faces more often compared to upright.

Experiment 2 was very similar to Experiment 1 but with the added condition of including the whole body with no background. The pattern of results produced by the whole-body condition were identical to the whole image condition results in both accuracy and RTs. Again this explanation of visual person representations rather than visual face representations helps to explain these results. Images seen in the inverted whole image condition again were typical of the kind of visual information stored in the visual person representation (a whole body), however as the image was inverted but the visual person representation contains primarily upright instances, it was harder to generalise and therefore a small number of errors were made. Very similarly, when participants saw images in the whole-body condition, these images contained the exact same information as the whole image condition in terms of what is processed by the visual person representation. Therefore the exact same process took place as in the whole image condition, and participants’ accuracy and RTs were the same. This experiment shows clearly that the interaction between image cropping and orientation observed for familiar identities across the present series of experiments is *not* to do with the presence of the background, as the presence or absence of a background made no difference to the results, but rather the presence or absence of the body, as the results are driven by the underlying visual person representations. When the cropped head was inverted, participants saw images which did not resemble normal visual input nor the content of the visual person representation, therefore, the visual person representation struggled to generalise and match the incoming image and participants’ accuracy fell sharply. For unfamiliar identities, again the explanation is the same as for Experiment 1, that when images were presented upright, participants’ accuracy was very high because they scanned their visual person representations and none of them matched the visual input, therefore participants quickly and accurately rejected them as not familiar. When images were inverted, this made the system more error prone and produced a small number of errors.

In Experiment 3, participants produced ceiling effects. This is because the task was very easy and required little cognitive effort. Participants first saw the famous identity's name followed by an image of them. Therefore the task was not 'do you recognise this face?' but 'Is this photograph of Michael McIntyre?' Straightforward recognition of familiar identities is a fast, effortless and automatic task, therefore by providing the name of the identity beforehand, the scan of familiar visual person representations is reduced to 1 and it is primed, making the task even easier, producing ceiling effects in all conditions. Main effects of orientation and slower RTs were found in all inverted conditions, again as inversion makes the visual processing system more error prone, however easy the task. Although any interaction between image type and orientation was masked by ceiling effects, a small trend in that direction is seen and can again be explained by the idea of a visual person, not face, representation. A smaller percentage of participants got an inversion effect in the whole image condition (33% compared to 44% in the cropped head condition) and overall participants got smaller inversion effects in the whole image condition. This is again because in the inverted cropped head condition, the stimuli are not visually similar to the information stored in the person representations as the body is missing and the person is inverted, the 2 factors combining together to produce a stimulus to which the person representation struggled to generalise. Similarly in Experiment 4, as discussed before, the task proved again to be easy for participants and suffered some experimental design flaws. As a result, participants sat at ceiling and any effects were masked. Although ceiling effects were found in Experiments 3 and 4, the RTs can still be examined. In both experiments, image cropping did not interact with orientation, producing similar increases in RTs for both inverted whole and cropped images. This could suggest that image cropping does not play a role in these experiments.

The results of Experiment 5 are also consistent with this explanation of visual person representations. For both the cropped and whole images of unfamiliar identities, the images were seen upright in an exposure phase. Face recognition is an automatic process and participants automatically accurately rejected the faces as unfamiliar and began forming a new representation. These visual person representations will have been very poor as they contained only 1 instance of the person, and so this would have been weak (prone to forgetting) and not robust (does not contain a range of within person variation and so cannot generalise to recognise a novel instance). Participants then saw the same image again at test and had to indicate if they had seen the image before or not. There was no difference between the cropped heads and the whole images in both accuracy and RTs because both visual person representations, whether they contain an instance of the cropped internal features or the whole person, were very weak and not robust representations. The improvement offered by the whole image over the cropped internal features is negligible in comparison to a familiar visual person representation for example. Therefore both representations were similarly weak and susceptible to forgetting, producing equivalent decreases in accuracy.

Accuracy levels in this experiment were lower in comparison to the accuracy for unfamiliar identities found in Experiments 1 and 2. This is likely mainly due to the difference in tasks. In Experiment 1 and 2, the task was to state if participants recognised an identity or not, which requires scanning familiar visual person representations which produced high accuracy, and this is the task we naturally do when we encounter faces in day-to-day life. In Experiment 5, the task was to say if participants remembered seeing the image earlier, which is not as natural a task as simple recognition, and it is harder as it requires additional steps / cognitive processes, for example recognition of the face and remembering if you have seen it before or not. Furthermore, the stimuli were more challenging in Experiment 5, as the cropped faces were cropped more harshly compared to previous experiments, removing most of the external features, also leading to the increased difficulty and lower accuracy.

The results of Experiment 6 can be explained in a similar way to the results of Experiment 5. For the unfamiliar identities, the task and explanation is the same as for Experiment 5. Performance for unfamiliar faces in Experiment 6 was lower compared to Experiment 5, and when images were inverted, they sat at floor level. This poorer performance can potentially be explained by the increased difficulty of the task. Firstly, the stimuli were cropped even more harshly – in Experiment 5 the oval crop was applied in relation to the chin, meaning the chin and therefore some of the external contour around the cheeks was visible. In this experiment, they were cropped more harshly to remove all external features, which fully eliminated any sense of the face and head shape. Also, the task was much harder as there were many more items to remember, due to familiar and unfamiliar identities being tested, and as the number of items to remember increases, accuracy decreases (Baddeley et al., 1975).

For the familiar identities, during the exposure phase, participants saw both the cropped internal features and the whole images upright. As face recognition is an automatic process, participants will have recognised the individuals. At test, in the upright conditions, participants will have again recognised the individuals and were then able to correctly say that they had seen them in the exposure phase. In the inverted conditions, for the whole images, even though the image was inverted and ~99% of instances in the visual person representation are upright, as the photograph contained all of the visual information processed by the visual person representation (e.g. the whole body) and familiar person representations contain lots of within person variation, participants were able to generalise their representations and recognise the identity. However, for the inverted cropped faces condition, the images seen at test did not at all resemble the previous instances of the individual stored in the visual person representation. The cropped faces contained only the internal features, whereas when we have seen this individual, they have their whole face and a body. This extreme cropping combined with inversion made the photos unrecognisable in the test phase (the visual person representation could not generalise to match it), and therefore participants could not accurately say they had seen the individual before, producing the large decrease in accuracy and increase in RTs for inverted cropped familiar faces.

As the accuracy is much lower and the reaction times were longer for the unfamiliar IDs compared to familiar IDs, this demonstrates that visual person representations are involved in the test phase. If they were not, then familiar and unfamiliar identities would produce similar levels of overall accuracy and RTs and a similar pattern of results, however, the greater accuracy and faster RTs in all familiar face conditions and the significant main effect of familiarity clearly demonstrates that remembering the faces is not independent of but dependent on the underlying visual person representations.

Overall, the results of this chapter clearly demonstrate that familiarity modulates inversion effects as measured by accuracy. For familiar identities, when images were cropped harshly, as had been done in most FIE research, very large inversion effects were found in the accuracy data; however, this effect was reduced when images were not cropped. This may be due to inverted cropped images of familiar faces not resembling the robust visual person representations (which contain the whole body), therefore inhibiting recognition. However, only Experiment 2 showed a modulating effect of image cropping in the familiar faces RT data, Experiments 1, 3, 4 and 6 did not. However, for unfamiliar identities, image cropping did not have an effect on inversion as measured by accuracy and RTs. This is because in recognition tasks, unfamiliar identities do not have underlying visual person representations, but instead the task required participants to scan their familiar visual person representations, which could accurately reject the unfamiliar faces as unfamiliar, regardless of the image cropping. In the memory tasks, there was no difference between the cropped faces and the whole images because in both conditions, participants formed small weak visual person representations. Both were such poor representations in contrast to a familiar visual person representation, that the difference between the cropped and whole image was almost non-existent and therefore produced equal inversion effects. The effects of inversion then are due to whether the incoming image resembles the underlying visual person representation or not and the strength / robustness of the representation. For familiar faces, the effect of image cropping appears to primarily affect accuracy scores rather than RTs. Only Experiment 2 found a modulating effect of image cropping on the inversion RTs. One possibility is that as upright familiar face perception producing ceiling accuracy, the RT data may reveal that image cropping does not play a large role in the FIE. However, as the data were not at ceiling in Experiment 6 and a large effect of image cropping was still found in the accuracy data, this may suggest that the effect of image cropping may only be strong enough to affect accuracy and not RTs. The use of familiar faces to study the face inversion effect has produced novel results and opens up an exciting new line of research.

# Chapter 4 – General Discussion

This thesis had 2 aims: to test whether top-down information plays a role in face learning (chapter 2), and to test if the face inversion effect is a stimulus-bound or genuine effect (chapter 3). Across these two different branches of face perception research, a common theme emerged: although face familiarity was not the initial subject of study, in both sets of experiments face familiarity played a key role, often producing larger effects than the originally studied effects.

### Overview of aims, findings and discussion – Chapter 2 Face Learning

In chapter 2, the role of top-down cues in face learning was investigated. The experiments were designed around the premise that observing within-person variation is crucial to face learning, and observing a greater range of variation gives rise to better recognition accuracy (Burton et al., 2016; Murphy et al., 2015; Ritchie & Burton, 2017). The aim of Experiment 1 was to run a successful replication study of Ritchie and Burton (2017) and to establish the advantage of observing greater within-person variation in preparation for Experiment 2. In this experiment, participants learnt to recognise faces by viewing low or high variation images which were blocked by identity with the ID’s name presented throughout (the same top-down cues used in the original paper). Following the learning phase, participants completed a matching task with new unseen images of the learnt identities as well with novel identities. The results of Experiment 1 did not replicate Ritchie and Burton (2017), there were no significant differences in the matching accuracy or RTs between identities learnt from low or high variation images.

Experiment 2 then sought to test if removing the top-down cues from face learning would reduce performance on high variation trials. In this experiment, new participants learnt to recognise the same identities without top-down information: identities were no longer blocked together and the name was removed from the screen. There was no significant difference in the accuracy between identities learnt from low or high variation images, and RTs were identical. The results of these experiments suggested that in the learning phase of Experiment 1, participants used the top-down cues to bind together the highly variable instances and began building face representations (Andrews et al., 2015; Schwartz & Yovel, 2016). In Experiment 2, in the absence of such cues, participants may have been less able to cohere together the highly variable images of identities, producing matching accuracy and reaction times which were very similar to the low variation condition.

However, as Experiment 1 was unsuccessful in replicating *significant* differences between identities learnt from low and high variation images, this makes the null effect of Experiment 2 less compelling. Therefore, the aim of Experiments 3 and 4 were to improve upon Experiments 1 and 2, in another attempt to find a significant difference between identities learnt from low and high variation identities when they were learnt with top-down information, and to see if removing such cues removes the high variation advantage. Experiments 3 and 4 were very similar to Experiments 1 and 2, however the experiment design was held constant, the statistical power was increased by increasing the number of trials and participants, and better stimuli were selected with less variable images used in the low variation condition. Experiment 3 found significantly higher accuracy and faster reaction times for identities learnt from high compared to low variation photographs, this time successfully replicating the advantage of observing greater within-person (Ritchie &

Burton, 2017). Furthermore, in Experiment 4 when top-down cues were moved from the learning phase, participants were no longer significantly more accurate at recognising identities learnt from highly variable images compared to low variation images, however, they were still matched significantly faster.

In Experiment 3, when identities were learnt with top-down cues, as the identities learnt from high variation images were matched significantly more accurately and faster compared to low variation identities, this suggests that participants were able to use the top-down cues (a background colour and name) to bind together the images into a face representation (Andrews et al., 2015; Schwartz & Yovel, 2016). For the identities learnt from low variation images, as the images look inherently similar, they were likely bound together into a face representation without the need of top-down cues. The differences in these underlying face representations constructed in the learning phase may have produced the differences in the matching task. The face representations of identities learnt from high variation images contained more within-person variation and were therefore better able to generalise and recognise novel images of the identity, producing the greater accuracy and speed in the matching task (Murphy et al., 2015). It is known that face representations of unfamiliar / familiarised faces (faces learnt from low variation images here) are less able to generalise and recognise novel instances of the identity (Liu et al., 2009), in this experiment likely leading to the poorer performance in comparison to the high variation identities.

However, the largest differences in Experiments 1 – 4, regardless of whether top-down cues were included in the learning phase or not, was the difference between novel and learnt identities. The difference in accuracy scores between novel identities and identities learnt from low variation images was larger than the difference between identities learnt from low and high variation images in all 4 experiments. This is particularly interesting as the pattern comes out in Experiments 2 and 4 when no top-down information was provided during face learning. This suggests that participants were able to tell the high and low variation images together in the learning phase quite well from bottom-up visual information alone. Therefore, this suggests that the role of top-down information in face learning is quite small, providing only a small benefit when viewing highly variable images.

These results also suggest that face representations may be built up faster and with greater ease than previously thought. When building a face representation, the observed face is initially completely unfamiliar, novel. The large differences between the processing of familiar and unfamiliar faces is now a very well-known and studied effect in the face processing literature (Bruce et al., 2001; Butcher & Lander, 2013; Clutterbuck & Johnston, 2004). For example, it is known that introducing even a small change to an unfamiliar face, such as different lighting and/or a different facial expression reduces recognition accuracy greatly, however such changes do not affect the recognition accuracy of familiar identities (Bruce, 1982). It is even argued that familiar and unfamiliar faces are processed *qualitatively* differently (Megreya & Burton, 2006). However, in Experiments 1-4 here, the difference in the accuracy and reaction times between unfamiliar faces (faces learnt from low variation images) and familiar faces (faces learnt from high variation images) is quite small and is not suggestive of a qualitative difference. This potentially large conflict with the literature can be explained by careful examination of the terms ‘unfamiliar’ and ‘familiar’. These terms are used often in the literature but exact definitions have not been pinned down. How much within-person variation is necessary or how long does a face need to be viewed before it is classed as familiar? There has been a move away from a binary definition of familiar/unfamiliar to a spectrum of familiarity (Kovács, 2020).

I would like to argue that faces which have either never been seen before or 1 photograph has been observed represent the lower end of the familiarity spectrum and can collectively be called unfamiliar, they are processed *qualitatively* differently from familiar faces, and faces in this category jump up to familiar with greater ease than previously thought. Identities learnt from low variation images in this experiment through to personally familiar faces represent the higher end of the familiarity spectrum and can collectively be called familiar. Within this range, faces are processed *quantitatively* differently from each other due to the differing size of the underlying face representation, and they are processed *qualitatively* differently from unfamiliar faces which either do not have underlying face representations (unseen faces), or face representations which contain only 1 instance of a face and therefore contains no within-person variation. The nature of a quantitative or qualitative difference in face processing is due to the presence or absence of a face representation which contains within-person variation. This proposed model of the spectrum of familiarity (see Figure 4.2) is laid out based on the model in Kovács (2020) (see Figure 4.1). It can also be conceived of graphically (see Figure 4.3).

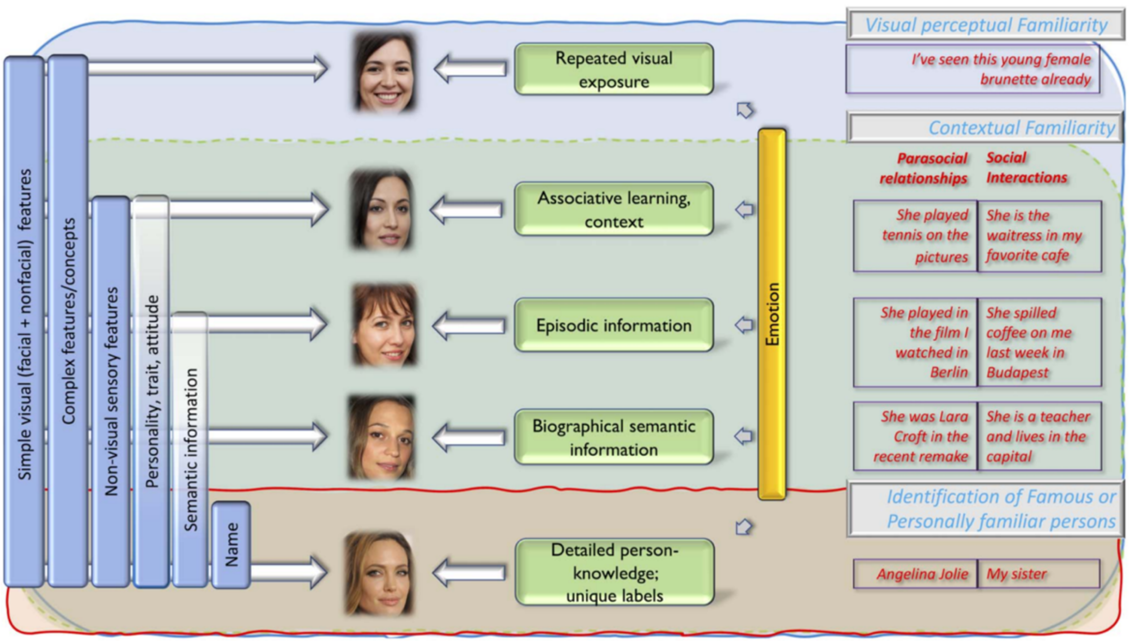


**Figure 4.1.** The model of face familiarity from Kovács (2020). This model is a more comprehensive account of increasing familiarity than the spectrum I am suggesting, however it does explicitly cover quant-/qualitative differences in processing.

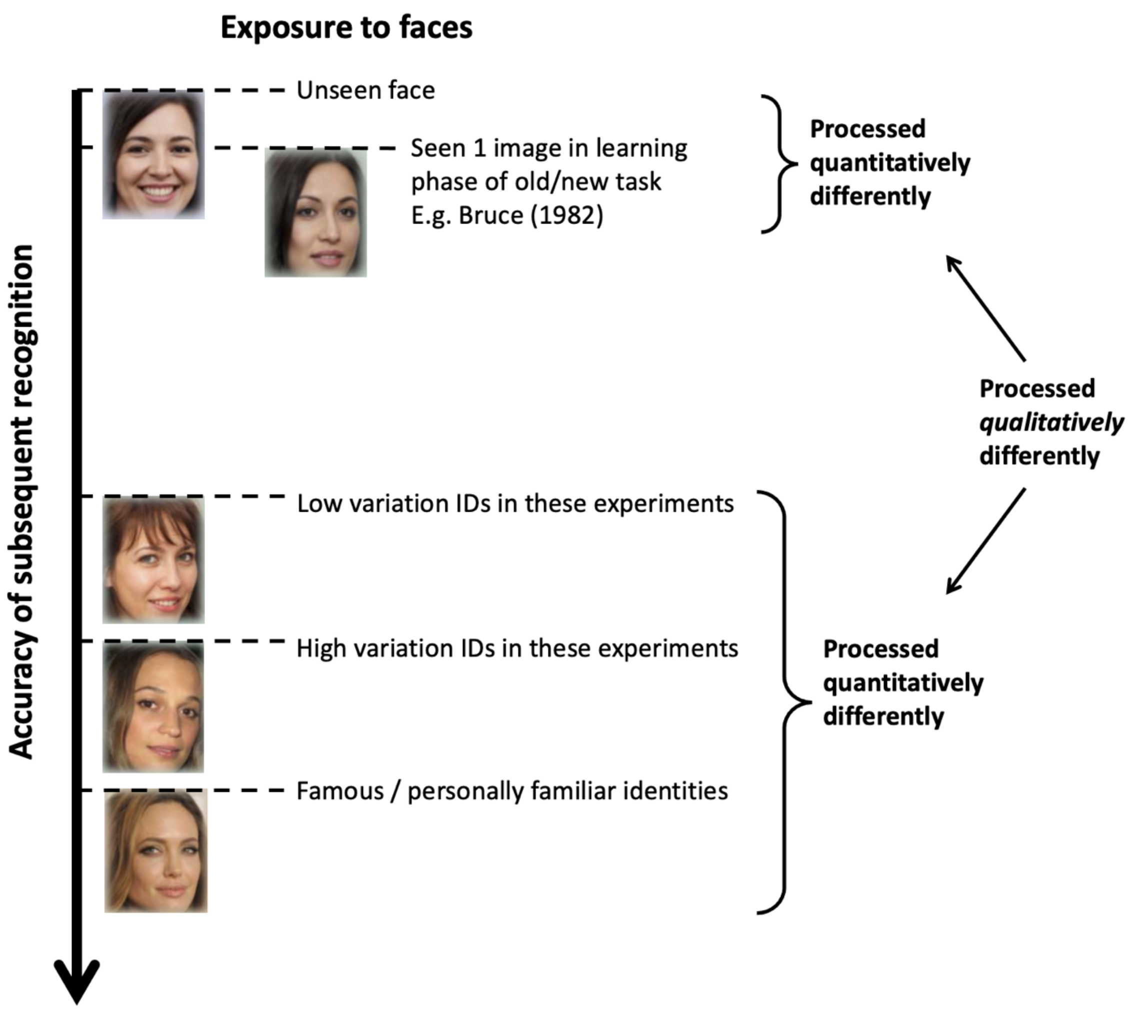



**Figure 4.2.** Proposed spectrum of face familiarity. Face images taken from Kovács (2020).

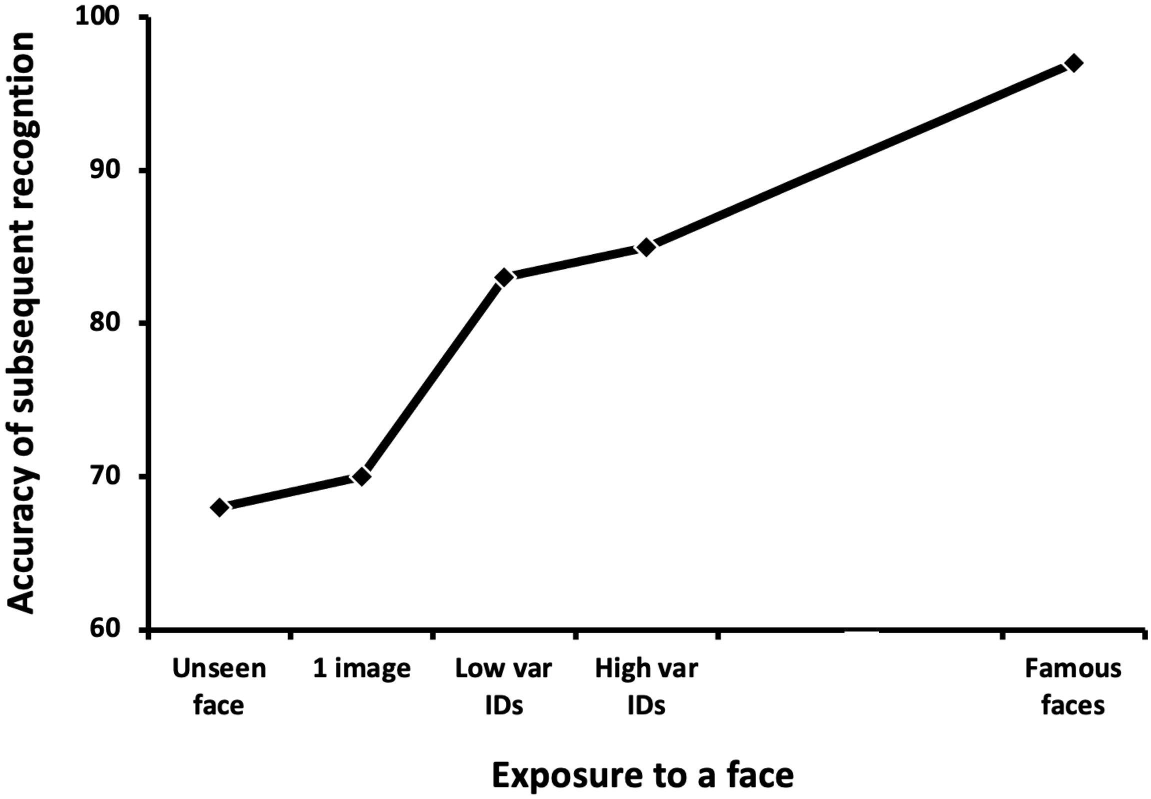



**Figure 4.3.** Proposed model of face recognition accuracy as a function of observed within-person variation. I predict this is a non-linear model, with a qualitative change in processing occurring when a face representation containing within-person variation is formed.

As shown in Figure 4.3, the steepest increase in subsequent recognition may be between where only 1 image of an identity has been seen before (no within-person variation) and between low variation IDs (minimal within-person variation). Unseen faces, e.g. faces never seen before, have been shown to be matched with approximately 70% accuracy (in Chapter 2 here, 77% in Experiment 1, 71% in Experiment 2, 72% in Experiment 3 and 69% in Experiment 4). Similarly, faces which have only been seen once may be recognised slightly more accurately, but in a very similar range (approx. 70-75% accuracy), for example in Chapter 3, unfamiliar faces seen once during an exposure phase were recognise with 76-78% accuracy in Experiment 5, and 56-66% accuracy in Experiment 6. There may then be a sharp increase in accuracy when within-person variation is seen. For example, in Chapter 2 when identities were learnt from 10 low variation images, identities were matched with 83% accuracy in Experiment 1 and 81% accuracy in Experiment 3. There is a small jump up as identities are learnt from 10 high variation images – in Experiment 1 IDs learnt from high variation images were matched with 86% accuracy and in Experiment 3 with 84% accuracy. Much further down the continuum would be famous or personally familiar faces which are recognised highly accurately (90%+). For example, in Chapter 3, in Experiment 1 upright famous faces were recognised with 92-98% accuracy.

Such a spectrum of face familiarity which incorporates changes in quant/qualitative processing could be a helpful addition to the literature as it offers clearer definitions of familiar and unfamiliar faces, and therefore clearer predictions can be made. For example, in Bruce et al. (1999), participants performed a matching task with 1 target image and an array of 10 faces to choose a match from – all of these faces were unfamiliar to the viewer and had never been seen before. On match trials, participants were on average 70% accurate. As these faces had never been seen before, they were truly unfamiliar and would be processed qualitatively differently from familiar faces. As participants had not seen these faces before, they did not have a face representation of the identity to use to match the target and array image to. Therefore, they might have used a different process, image matching strategies which are closely bound to the images, in order to perform the task. Unfamiliar face matching is often conceived of as highly error prone and the emphasis is placed on the approx.

30% errors participants make (Bruce et al., 1999), however, it is also important to note that 70% accuracy is 20% above chance level, so it is not that participants are *unable* to match unfamiliar faces, but that they are less accurate.

Furthermore, faces which have only been seen once would also be classed as unfamiliar, with the same perceptual processes used as identities which have never been seen before. In a matching task with unseen (unfamiliar) faces, the participant views two images, the target image and the matching image in the array; this is similar to a memory task where a participant views two images of the identity, one in the exposure phase and one at test. For example, in Experiment 1 of Bruce (1982), an old-new recognition task was used and in the exposure phase, participants saw 1 photo of each unfamiliar identity. According to the spectrum of face familiarity put forth here, these faces would be classed as unfamiliar as no within-person variation had been observed. When 1 change was made to the face at test (either a different viewpoint or expression), recognition accuracy decreased by 13.5% compared to viewing the same image (from 89.6% accuracy for the same image to 76.0% for 1 change). Furthermore, when 2 changes were made, accuracy decreased by a further 15.5% (from 76.0% for 1 change to 60.5% for 2 changes). This fits with the model put forth here, as only 1 image had been viewed during the exposure phase, participants had not seen within-person variation and had begun to create a face representation not capable of accurately generalising to novel instances of the same identity. The incremental decreases in accuracy as further changes are made to the face demonstrate that the ability of the weak face representation to generalise is closely bound to the 1 image seen in the exposure phase. However, in Experiment 2 of Bruce (1982), participants familiar with the faces were also tested on the old-new recognition task. Their performance was unaffected by 2 changes being made to the test image.

As the faces were personally familiar to the participants, they had observed a wide range of within-person variation of the identities and formed robust stable face representations, capable of generalising accurately to novel images (Bruce, 1994). Therefore, these images would sit at the higher end of the spectrum put forth here and are processed qualitatively differently from unfamiliar faces, relying on a robust face representation rather than image matching strategies.

This spectrum of familiarity, defined by the amount of within-person variation observed, helps to explain the pattern of the chapter results. For example, identities learnt from low variation images might be considered unfamiliar identities, however, as within-person variation was observed, according to the proposed spectrum they would therefore be considered familiar identities. In the matching task, the images were perceived in a qualitatively different way to the novel identities (using the face representation built up in the learning phase compared to image matching strategies), and therefore performance was significantly more accurate compared to novel identities, on average across the 4 experiments matching performance was 9.01% more accurate. For the identities learnt from high variation images, these would also be considered familiar identities, and were processed in a *quantitatively* different way to identities learnt from low variation images, on average across the 4 experiments matching performance was 2.45% more accurate. Both low and high variation IDs formed face representations, and the novel images in the matching task were matched again these representations; however, for identities learnt from high variation images, more within-person variation had been observed, but the underlying process used was the same as identities learnt from low variation images, hence a quantitative difference.

It has been shown that observing a greater range of within-person variation leads to better subsequent recognition accuracy (Ritchie & Burton, 2017), however, from these results it seems that exposure to *any* level of within-person variation, even 10 low variation photographs, leads to much improved recognition accuracy. In these 4 experiments the biggest differences do not lie in the amount of within-person variation observed (low vs high variation IDs), but in whether any within-person variation is observed or not (novel vs learnt IDs). This is reflected in the proposed idea that novel identities are unfamiliar and both low and high variation identities are familiar identities, which are processed quantitatively differently from each other and qualitatively differently from novel identities. For example, in Experiments 2 and 4 here it was predicted that when high variation identities were learnt *without* top-down cues that they would be matched with similar levels of accuracy as identities learnt from low variation images, however, they trend towards being recognised more accurately, and they are matched significantly more accurately than novel identities. This is surprising as recognition of unfamiliar identities suffers with changes in appearance (Bruce, 1982) and unfamiliar viewers struggle to tolerate within-person variation (Andrews et al., 2015; Jenkins et al., 2011). This suggests that faces may be learnt with greater ease than previously thought.

I predict that when any amount of within-person variation is viewed, the minimum being two photos of an identity, that matching accuracy will start to increase rapidly, as the process changes from image matching strategies to a face representation generalising to accept a novel image (Johnston & Edmonds, 2009). Future experiments could test this idea of a non-linear function with an old-new memory task as in Experiments 1-4, and also testing identities learnt from 1, 2, and 6 images for example, to test the slope of the increase in accuracy. I predict that identities learnt from 1 image will be matched with similar accuracy as novel identities. Identities learnt from 10 images will be matched more accurately than identities learnt from 6 images, which will be matched more accurately than identities learnt from 2 images, which will be matched more accurately than identities learnt from 1 image. I predict that the slope will increase sharply somewhere between identities learnt from 1 and 10 images, as a qualitative processing change occurs and the process shifts from basic image matching strategies to recognition via face representations.

Initially, the key results of this chapter, that faces are learnt rapidly, might seem at odds with the eye witness testimony literature. There is now a vast body of evidence which suggests that eye witness testimony is unreliable and yet is given considerable weight in court (Loftus, 1980). The results from Experiments 1-4 suggest that faces can be learnt quickly leading to accurate recognition, yet the eyewitness testimony literature shows that subsequent recognition is not accurate (Deffenbacher et al., 2008). Two large differences may help to explain the discrepancy. Firstly, viewing a real life crime is a very different and stressful situation compared to taking part in a laboratory experiment. For example, when witnessing a violent crime, weapon focus may make viewers fixate more often and for longer at the weapon compared to the face (Loftus et al., 1987). Such factors are not present in face learning laboratory experiments and furthermore, in experiments, participants are instructed to pay attention to the experiment with minimal distractions.

Secondly, recognition tests in experiments tend to take place very soon after viewing a face (either immediately as in the present experiments or after a very short delay), whereas identification of a suspect might take place days or even weeks after witnessing a crime. It has been shown that delaying the identification of a face to 1 day later compared to 1 hour later leads to a large decrease in accuracy, and accuracy continues to decrease further after 1 week and again after 2 weeks (Wixted & Ebbesen, 1991). Furthermore, a meta-analysis found that the longer the retention interval, the lower the recognition accuracy, in both face recognition and eyewitness identification studies (Deffenbacher et al., 2008). Taken together, the vast differences between the experience of viewing a face in a real-life crime or an experiment, and being tested on identification minutes or days later may lead to the lower recognition accuracy observed for eyewitnesses compared to participants in laboratory experiments.

**Overview of aims, findings and discussion – Chapter 3 Face Inversion**

In chapter 3, the role of image cropping on the face inversion effect was investigated. The face inversion effect (FIE) is the disproportionately larger drop in accuracy when recognising inverted faces compared to other stimuli (Yin, 1969) and is typically studied using cropped images of faces (Hadad et al., 2019; Hole, 1994). Experiment 1 aimed to test if using whole images (not cropped around the face but rather including the body and background) would reduce the face inversion effect. Participants were asked to indicate if they recognised photographs of familiar faces and buildings which were either presented upright or inverted, and as whole images or cropped around the external features. For cropped images, the face inversion effect was replicated, famous faces suffered much larger decreases in recognition accuracy compared to familiar buildings (Diamond & Carey, 1986; Yin, 1969). However, for whole images, familiar faces and buildings suffered similar sized inversion effects. Unfamiliar items were included in this experiment for participants to respond “no” to, they were not the focus of interest, however, unfamiliar faces showed a different pattern to the familiar faces, and suffered similar small decreases in accuracy whether they were shown as cropped or whole images, which was similar to the small decrease in accuracy seen for inverted whole familiar faces. As in chapter 2, familiarity was not the initial focus of research, but again large, interesting findings were observed, which later drove the direction of the chapter.

The results of Experiment 1 possibly suggested that context cues in the whole images (e.g. the backgrounds) cued recognition, which partially overcame the effect of inversion (Gruppuso et al., 2007). However, an alternative explanation of the results was that cropped inverted images looked too dissimilar to the stored face representation and could not be recognised, but the whole images were similar to the face representation and were therefore recognised. Experiment 2 aimed to pull these two ideas apart by creating a similar recognition experiment with an additional condition with the background cropped out but the whole person (head and body) intact. Participants again suffered large decreases in accuracy for inverted cropped images of famous faces, however, both whole body and whole images suffered similar smaller decreases in accuracy and identical increases in RTs. This suggests that the background played no role in cuing identity, as there was no difference between a whole body with the background included or removed, but rather the difference lay between images of whole bodies and cropped heads. Therefore the results from Experiment 2 suggested that removing the body caused the large decrease in accuracy. This may be because familiar faces are recognised by their underlying face representations. As the identities were familiar, the participants had observed within-person variation of the famous faces before on TV, in films, on social media etc. and built up face representations. New instances of the identity can be recognised, e.g. seeing an actor in a new film, as the face representation is able to generalise and the incoming image is accurately matched and then recognised. However, in this experiment, the body is removed and the image inverted, which does not represent the previous visual experience participants had with the famous identity, and therefore the representation cannot generalise and accurately recognise the image.

Experiment 3 aimed to test if the reduced inversion effect for whole naturalistic images could be replicated across different tasks. A name verification task was used, whereby participants saw a name followed by an image of a famous identity either upright or inverted, and either as a cropped head or a whole image, and participants had to indicate if the name and image were of the same person or not. Ceiling effects masked any clear patterns in the data as the task was quite easy. As the name appeared first, this may have reduced the search across face representations down to just one, and primed the representation. Furthermore, as each image was presented twice in the experiment, participants may have been able to use that information to correctly guess the answer upon the second presentation. This idea was confirmed by a t-test showing participants were more accurate on the second presentation. This was a weakness of this experiment which could be improved by using 2 different images of the famous identities.

In Experiment 4, another attempt was made at finding the reduced inversion effect for whole naturalistic images using a different task. Yin's (1969) original method was used, an old-new memory task. Participants saw images of famous identities in an exposure phase, then at test indicated if they thought they saw images of the targets and distractors in the exposure phase or not. Any pattern in the results were again masked by ceiling effects due to the ease of the task and some experimental design issues. Therefore, Experiment 5 was run to address these issues. In an improved experiment, participants completed an old-new task with unfamiliar faces.

The results revealed equivalent decreases in accuracy and increases in RTs for both inverted cropped and whole images. This result was in stark contrast to the large difference between the inversion effects for cropped and whole images of famous faces found in Experiment 1. There were two differences between Experiments 1 and 5 which may have explained the differences in results: Experiment 1 focussed on familiar faces whereas Experiment 5 used unfamiliar faces, and Experiment 1 was a recognition task and Experiment 5 was a memory task. Therefore, Experiment 6 aimed to test if familiarity modulated the inversion effect by holding the experiment design constant and directly testing familiar and unfamiliar faces.

In the final experiment of this thesis, an old-new task was used again, testing both unfamiliar and famous faces. Participants viewed the faces in an exposure phase, then at test the target identities were mixed in with distractors and participants indicated if they had seen them before or not. The pattern found in Experiments 1 and 2 for famous faces was replicated again in this experiment: for cropped famous faces, inversion caused a large decrease in accuracy in comparison to whole images which showed a much reduced effect. However, for unfamiliar faces both cropped and whole images suffered reductions in accuracy taking performance down to chance when inverted, however, the drop was larger for whole images. The larger decreases for whole images of unfamiliar faces was not seen in any previous experiment, and for famous faces the opposite pattern is found, suggesting that the pattern was masked by floor effects. Therefore, the results of Experiment 6 and the chapter as a whole showed that inversion effects are modulated by familiarity: that for familiar faces, cropped faces suffer large inversion effects whereas whole images do not, and that inversion effects with unfamiliar faces are not affected by image cropping.

As discussed above, the pattern of results from Chapter 3 may be explained by the cropped images not resembling how we see faces in day-to-day life (e.g. the body cropped out). Therefore, for familiar faces, as the encoded image is missing the body, it is encoded further away from the face representation than a whole uncropped image would be, and therefore recognition is impaired. These findings and explanation fit with Brandman and Yovel (2012). In their experiments, participants completed a sequential matching task upright or inverted with either a head with a face, a faceless whole body, a headless body, a faceless head, a faceless head with the upper body, or the body from behind. They found that a head with a face, a faceless whole body or a faceless head with the upper body all suffered large inversion effects. In a face detection task when stimuli were shown for only 27ms, they found that faceless whole bodies and faceless heads with the upper body were perceived as containing a face more than a faceless head without a body. They also found that a faceless whole body and a head with a face were perceived similarly as having a face. This suggests that the body is part of face perception, and even seeing just the upper body generates the percept of a face.

To test whether their results with faceless image cropping is also modulated by familiarity, as found in chapter 3 here, a similar experiment combining the methods of the present experiments and those of Brandman and Yovel (2012) could be conducted. Whole images (including the body and background) and cropped images (removing all of the body and background), like those used in chapter 3, could be used but with the faces removed. A recognition task like in Experiment 1 and 2, or a sequential matching task like that of Bradman and Yovel (2012) could be used with familiar and unfamiliar faces presented upright or inverted. Bradman and Yovel (2012) tested unfamiliar faces in their experiment and found a small inversion effect for faceless heads and a larger inversion effect for faceless whole bodies. It could be hypothesized in this future experiment idea that familiar faceless heads would also have a small decrease in accuracy when inverted, because there is no face or body so the visual person representation is not activated. It, like a faceless unfamiliar head, may be processed as a complex visual pattern, not a face. For the faceless whole images of familiar identities, as the body is included, it may activate the visual face (or person) representations. However, as the image is missing the face and is inverted, it may be encoded further away from the face representation than the previously observed instances of the identity and therefore produce similar large inversion effects as found in the present chapter.

Furthermore, literature on the FIE was covered in chapter 1 including multiple potential theories as to the cause of the effect: the holistic theory, the configural processing theory and the expertise hypothesis. Studies supporting these theories have tended to use unfamiliar faces and they are almost always cropped images. Future research could re-examine some key methodologies, such as the part-whole task (Tanaka & Farah, 1993), altering the second order configuration or changing the features of faces (Freire et al., 2000), or comparisons of experts and novices with objects of expertise and faces (Campbell & Tanaka, 2018; Diamond & Carey, 1986), but comparing cropped vs whole images for unfamiliar vs familiar stimuli. Whilst none of these theories were directly tested in the present thesis, the results of chapter 3 do not support these theories as explanations for the FIE. This is because whether holistic processing, configural processing of faces, or configural processing of objects of expertise was carried out on the faces in the present experiments, the same kind of processing should have been activated by upright images and inhibited by inversion, regardless of image cropping. However, using whole images of familiar faces drastically reduced the effect on inversion in the accuracy data. Therefore, it could be hypothesized that if the key studies mentioned above were rerun using cropped and whole images with familiar and unfamiliar faces, similar results might be obtained for the unfamiliar faces, however, effects may be reduced when using whole images of familiar identities.

Similarly to the face learning experiments in chapter 2, the initial focus of chapter 3 was not familiarity, yet face familiarity again proved to show some of the largest effects across the experiments. Earlier in this chapter, based on the results of Experiments 1 – 4 in chapter 2, I suggested that faces may be processed qualitatively differently depending on if they have an underlying face representation containing observed within-person variation or not. Completely novel faces or faces only seen once would be classed as unfamiliar, and faces seen more than once before would be classed as familiar, and these two categories are processed qualitatively differently from each other. These two very different categories of faces lead to different behavioural patterns, which can offer an explanation of the results found in these 6 inversion experiments.

As discussed in chapter 3, the nature of such face representations are perhaps like the multidimensional face space model put forth by Valentine (1991). Such a space is made up of many dimensions on which a face could vary, for example, orientation, viewpoint, brightness etc. In Valentine's (1991) model, a face is represented by one point in multidimensional space, however, I suggest that each instance of a face is encoded in multidimensional space and the region represents the identity and the centre the average of the identity. The amount of within-person variation observed, the amount of visual information in the representation, is the basis for the quantitative differences between familiar faces – the more within-person variation observed, the better the accuracy of subsequent recognition. These representations are built up by observing within-person variation. According to Valentine (1991), first a face in encoded into face space, then secondly a decision process occurs to determine if the encoded point matches an already known face or not.

*Encoding a face into face space*

I propose that the distribution for each dimension in face space is not necessarily a normal distribution, but instead the shape of the distribution for a given dimension is dependent on previous visual experience. Taking the two dimensions manipulated in chapter 3, orientation and amount of body visible, it could be suggested that such dimensions do not share the same distribution. For example, faces are extremely often observed upright in day to day life, and therefore the vast majority of instances encoded in face space lie at the upright end of the orientation dimension (see Figure 4.4). Faces are sometimes seen at a slight angle when someone tilts their head either to the left or the right, and therefore some instances will lie slightly further away from upright. Due to the biological constraints of neck length and the presence of shoulders blocking the head's path, heads do not tilt beyond approximately 90 degrees, and are rarely seen inverted, therefore very few instances are encoded at the 180° end of the dimension. Therefore, prior visual exposure with a face may produce an orientation dimension with an exponential decrease distribution, as show in Figure 4.4.

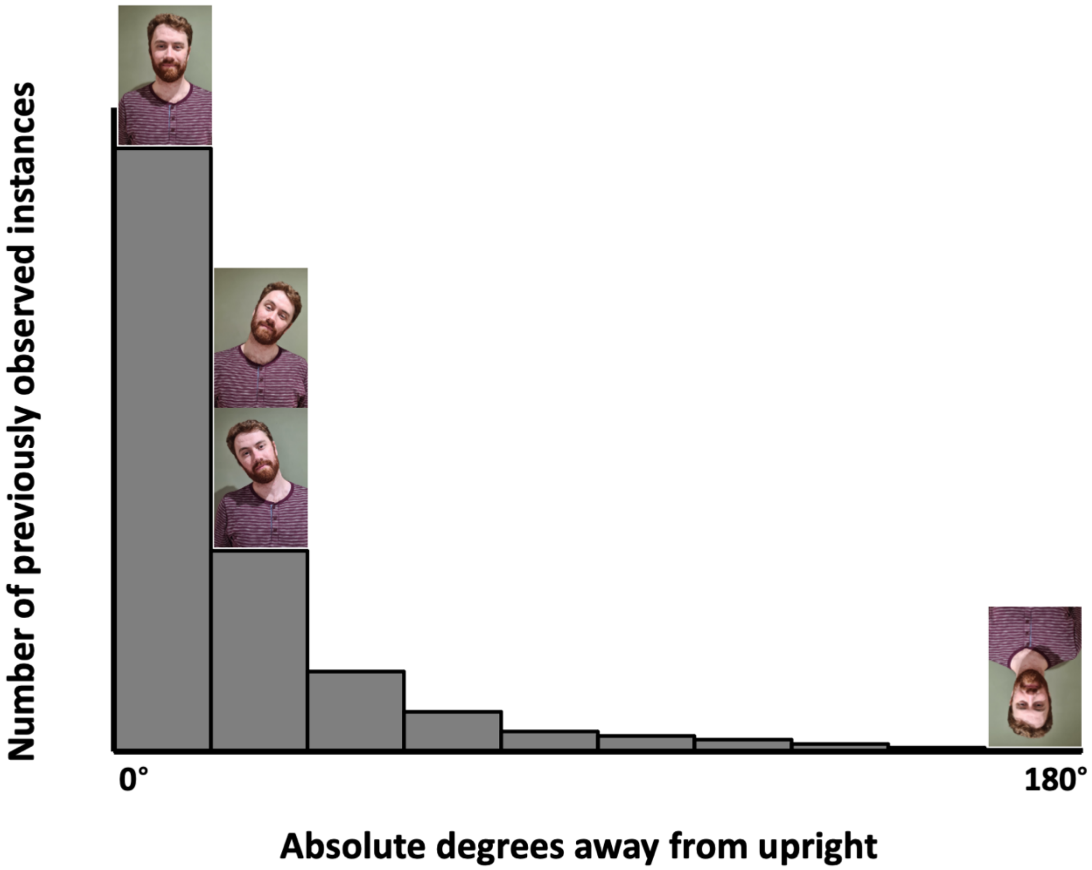



**Figure 4.4.** The orientation dimension of face space may have an exponential decrease distribution due to the high frequency of seeing upright faces and the very low frequency of seeing inverted faces.

Furthermore, the amount of a person we see varies and may be another dimension of face space. I proposed in chapter 3 that when we observe faces in day to day life, we do not just see a cropped face/head as in laboratory experiments, but we see the internal and external features of the face and the body, and all of this information forms part of the representation. Therefore face representations could instead be thought of as whole body representations or visual person representations. When seated at a work desk, coffee shop or taking a photo, we often see roughly half of the person. Even when talking stood up without any occlusions to the body, we attend to the face and due to our foveal/peripheral vision, we see the face clearly and approximately half of the body (Foi & Boracchi, 2016; Stewart et al., 2020). Therefore, this partial view may form the majority of observed instances. We see a whole person infrequently, for example only for a few seconds as they approach at a distance, then the majority of the time is spent in closer proximity with the lower half of the body out of view. Furthermore, we also see a heavily occluded person only infrequently, for example when covered by a bed sheet or placing their face through a photo stand-in. Therefore, the amount of person visible dimension may have a normal distribution, as shown in Figure 4.5.

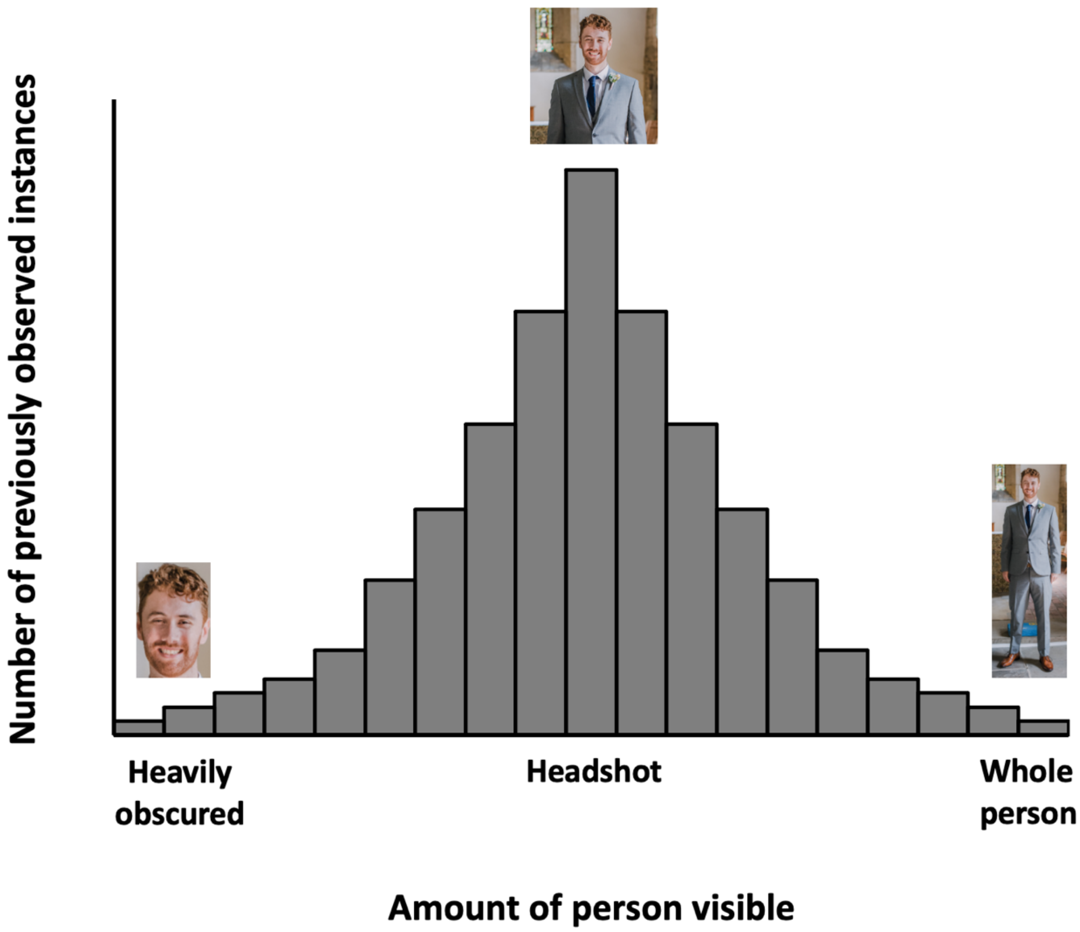



**Figure 4.5.** The amount of person visible dimension of face space may have a normal distribution.

Figure 4.6 shows these two dimensions of face space, orientation on the y axis and amount of person visible on the x axis. The blue dots represent previously observed instances of the identity which have been encoded into face space. The instances were encoded into face space based on their visual properties falling at a specific point on a given dimension. Due to the high frequency of observing the identity upright, many points are clustered near the 0° end of the orientation dimension, and due to the high frequency of observing the identity at a medium distance providing a view of the face and upper body, many points are clustered around the middle of the amount of person visible dimension.

When a face is observed, it is encoded into face space based on where it sits on various dimensions. It is a visual process and at this stage recognition has not yet occurred. According to Valentine (1991), faces are encoded with a degree of error, and inversion increases this error. I agree that inversion increases the encoding error, as when unfamiliar faces were inverted in all experiments in chapter 3, accuracy decreased, meaning participants were more often incorrectly responding that they recognised the individual. This therefore suggests that inversion does not simply impair recognition, as supposed recognition rates increased, but rather it increases errors.

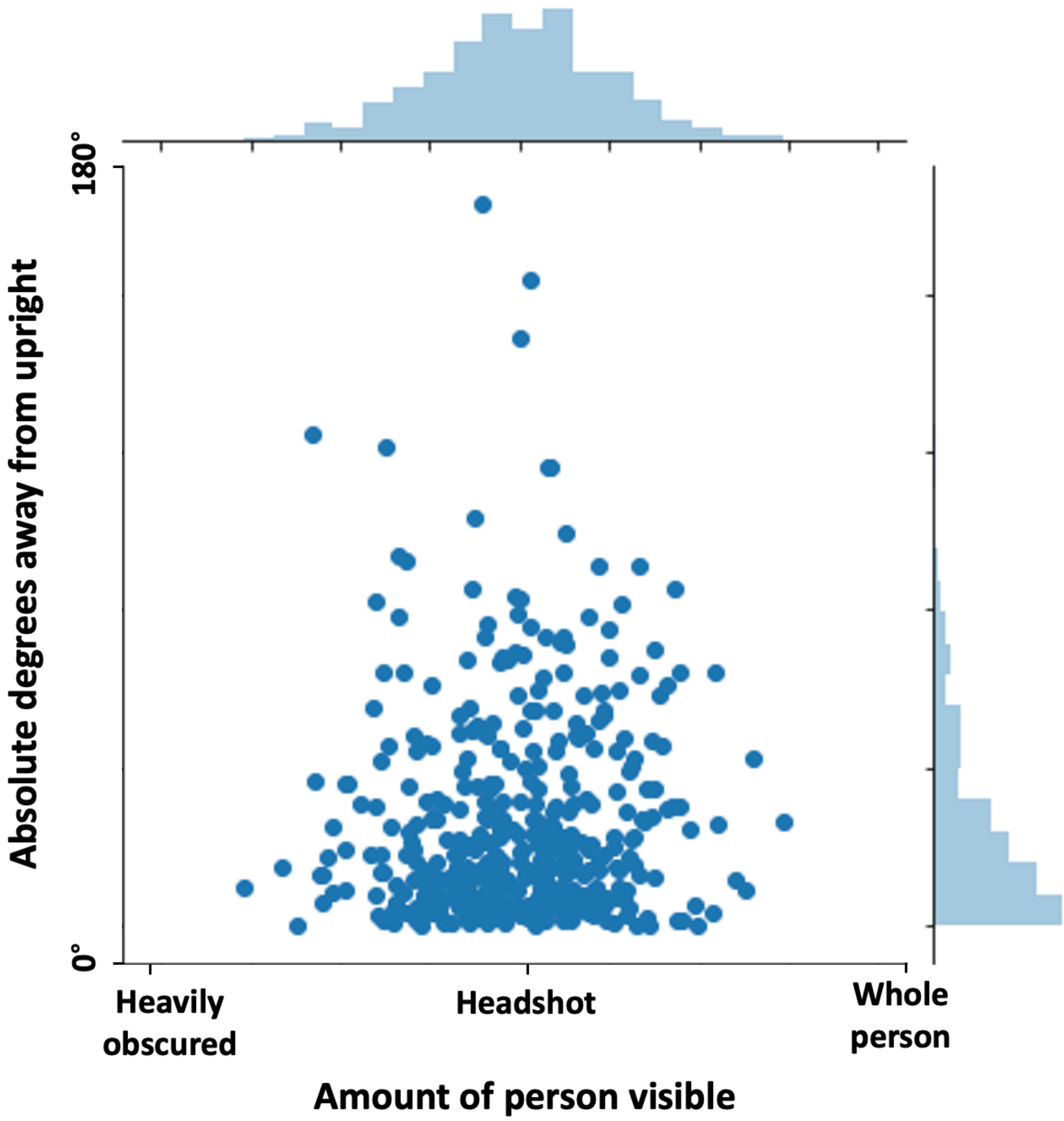
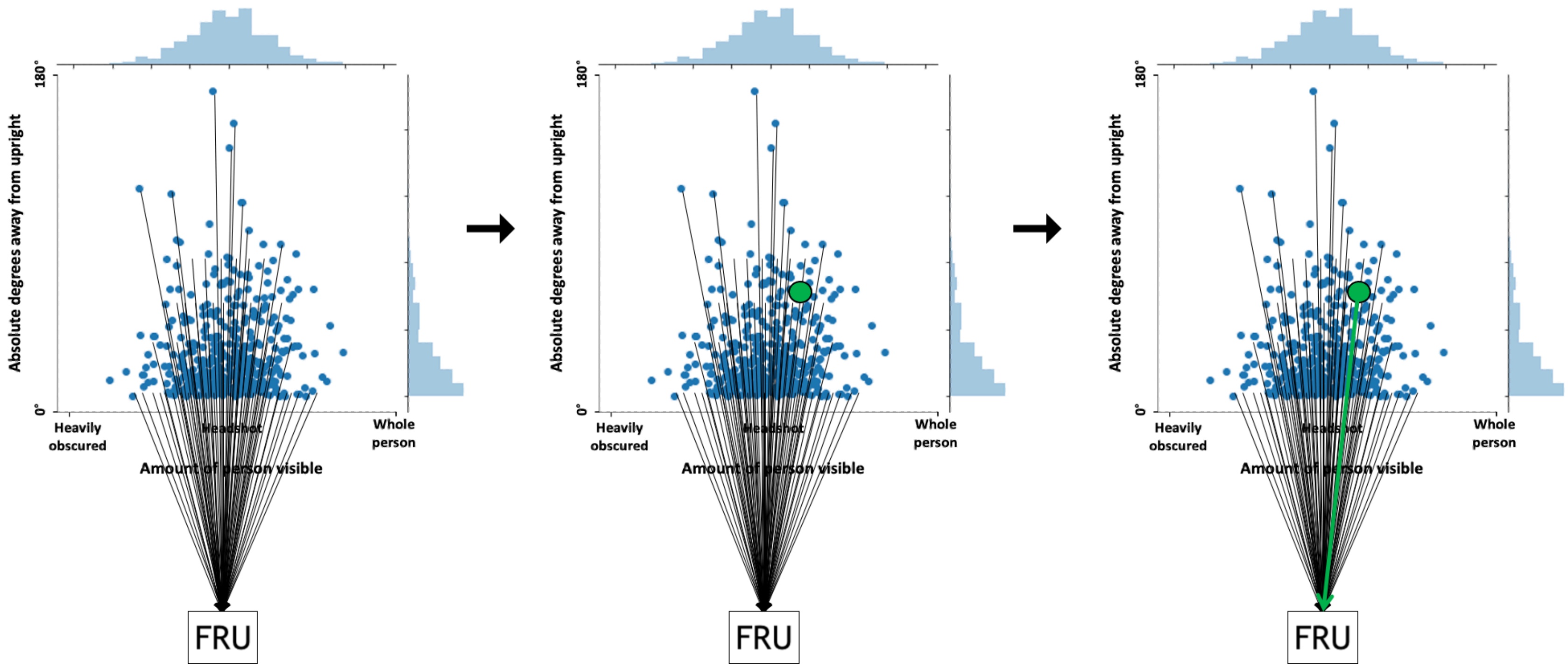



**Figure 4.6.** Two dimensions of face space, orientation and amount of person visible. The blue dots represent previously viewed instances of an identity which have been encoded into face space. The pattern of encoded instances is due to observing faces upright and viewing the face and upper body frequently. This graph shows 400 instances taken randomly from a normal distribution (upper x axis, amount of person visible) and an exponential decrease distribution (right-hand side y axis, absolute degrees away from upright).

*Decision process*

According to Valentine's model (1991), once a face has been encoded into face space, a decision process occurs which determines if the newly encoded face matches an already known face or not. He suggests in his exemplar-based model, that 3 factors affect the decision process: an estimate of the encoding error, the distance between the closest known face, and the distance between the nearest neighbour (a previously seen unfamiliar face one would not recognise). It is hard to pin down the exact mechanism which occurs, but it may involve calculating the distance from the encoded face to the nearest previously encoded face, and if the distance falls above or below a certain threshold, it may be classed as the same or a different person. It could perhaps be that the distance from the encoded face to the nearest face along each dimension is calculated, and if a high enough proportion of nearby faces belong to the same identity, it is also classed as that identity.

Whatever the exact mechanism is, I would suggest a key feature is the proximity from the encoded instance to nearby previously encoded instances. Once a new instance of the face has been encoded near other instances, it has only been encoded visually, it has not yet been recognised. A decision process then takes place involving the calculation of distances to nearby instances. Once a face has been accepted as the same as the nearby previously encoded instances, recognition has still not taken place yet. It is simply that the seen face is the same face as the other similar-looking nearby instances of the face, the identity is not yet known. I suggest that the previously encoded instances of the face project to a face recognition unit (FRU), as in Bruce and Young's (1986) influential face recognition model. Once the FRU is activated, then face recognition is achieved. Therefore, when a new instance has been encoded, and accepted as the same face as nearby instances, it then also projects to the FRU and is then recognised, as shown in Figure 4.7. When a new instance of a face in observed, the probability of it being correctly recognise is determined by its proximity to previously encoded instances, as shown in Figure 4.8, with darker regions representing a higher probability and lighter regions a lower probability.




**Stage 1:** Previously encoded instances formed many-to-one links to a face recognition unit (FRU).

**Stage 2:** A new instances is visually encoded (green dot).
**Stage 3:** A decision process based on proximity to nearby instances decides this is the same face.

**Stage 4:** Recognition. As the other instances project to a specific FRU, the new accepted instance also projects to the same FRU and the face is recognised.

**Figure 4.7.** Stage 1, instances in face space project to a face recognition unit (FRU). Stage 2, a new instance of the identity is seen and encoded into face space. Stage 3, a decision making process determines the newly encoded instance is the same face as the nearby instances. Stage 4, the new instance also projects to the FRU and the face is recognised.

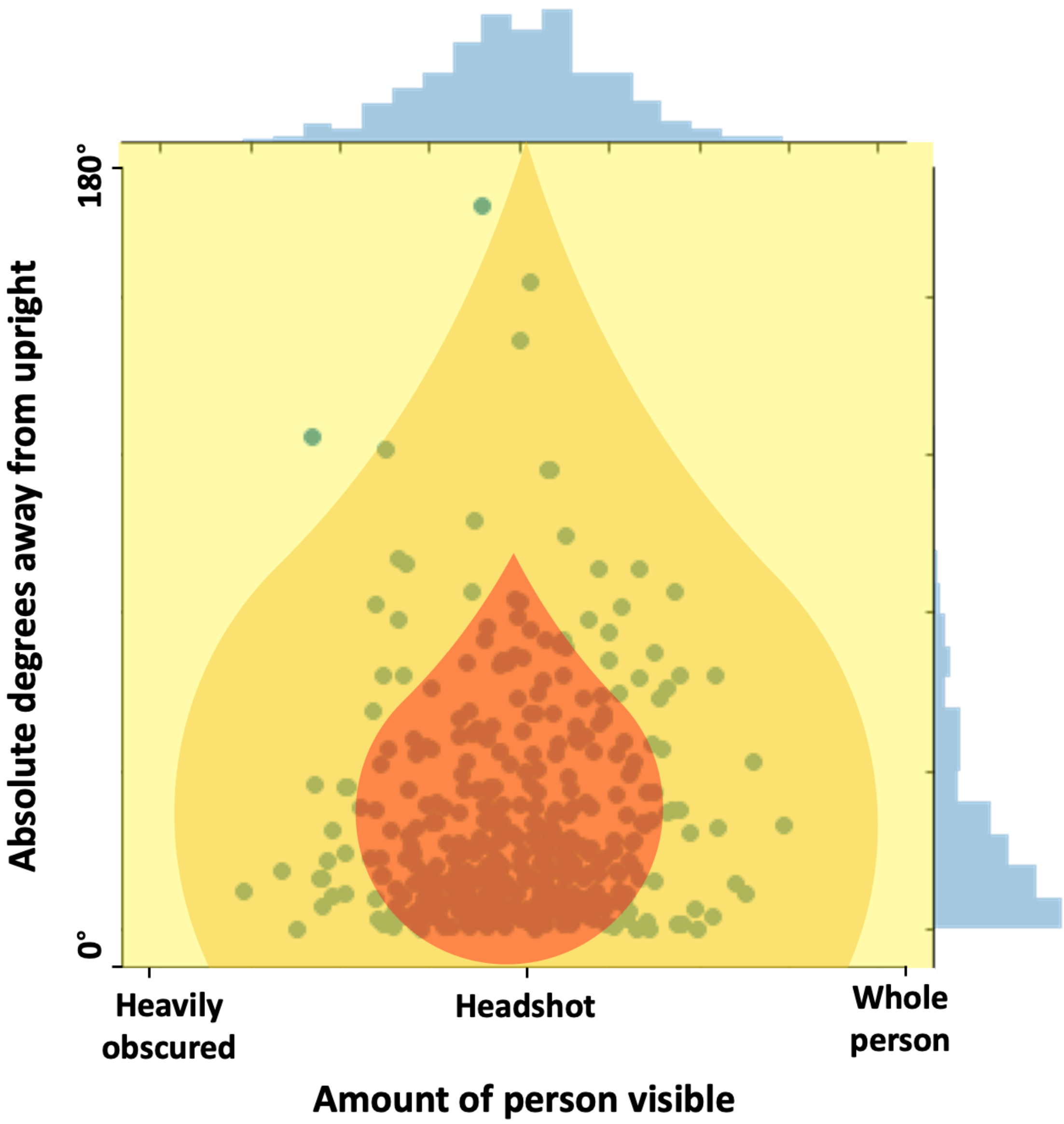



**Figure 4.8.** The probability of a newly encoded instance being encoded and correctly recognised is determined by the previously encoded instances. Darker colours

represent a higher chance that if a face is encoded within that region, it will be recognised.

Such a model of face space helps to explain the results of chapter 3. Figure 4.9 shows two dimensions of face space again, amount of person visible and absolute degrees away from upright. In Experiments 1 and 2 which used a straightforward recognition task, when famous faces were viewed as upright whole naturalistic images (red point in Figure 4.9), the image was encoded into face space based on the visual properties of being upright and including approx. half the body. As the encoded instance is surrounded by many close previously encoded instances of the same face, the face is matched to the representation, and then projects to the FRU and accurate recognition is achieved. When the whole image was inverted (green point in Figure 4.9), the image was encoded at the same location on the amount of person visible dimension, but at the opposite end of the orientation dimension. The encoded image was still surrounded by some previously encoded instances, though less and perhaps a little further away, so the image was still matched and recognised, but with slightly less accuracy and/or slower RTs. For the cropped images of famous faces, when upright (yellow point in Figure 4.9), the face is encoded at the same point on the orientation dimension as the whole upright face, but is further away on the amount of person visible dimension. Like with the inverted whole image, the encoded image was again surrounded by some previously encoded instances, though less and perhaps a little further away, so the image was still matched and recognised, but with slightly less accuracy. The cropped inverted head however, (purple point in Figure 4.9) was encoded at the far ends of both the amount of person visible and orientation dimensions. As such stimuli do not resemble our normal day to day visual experience, very few previous instances surround the encoded image. As previously encoded instances were further away, the image was matched less often and recognition accuracy decreased.

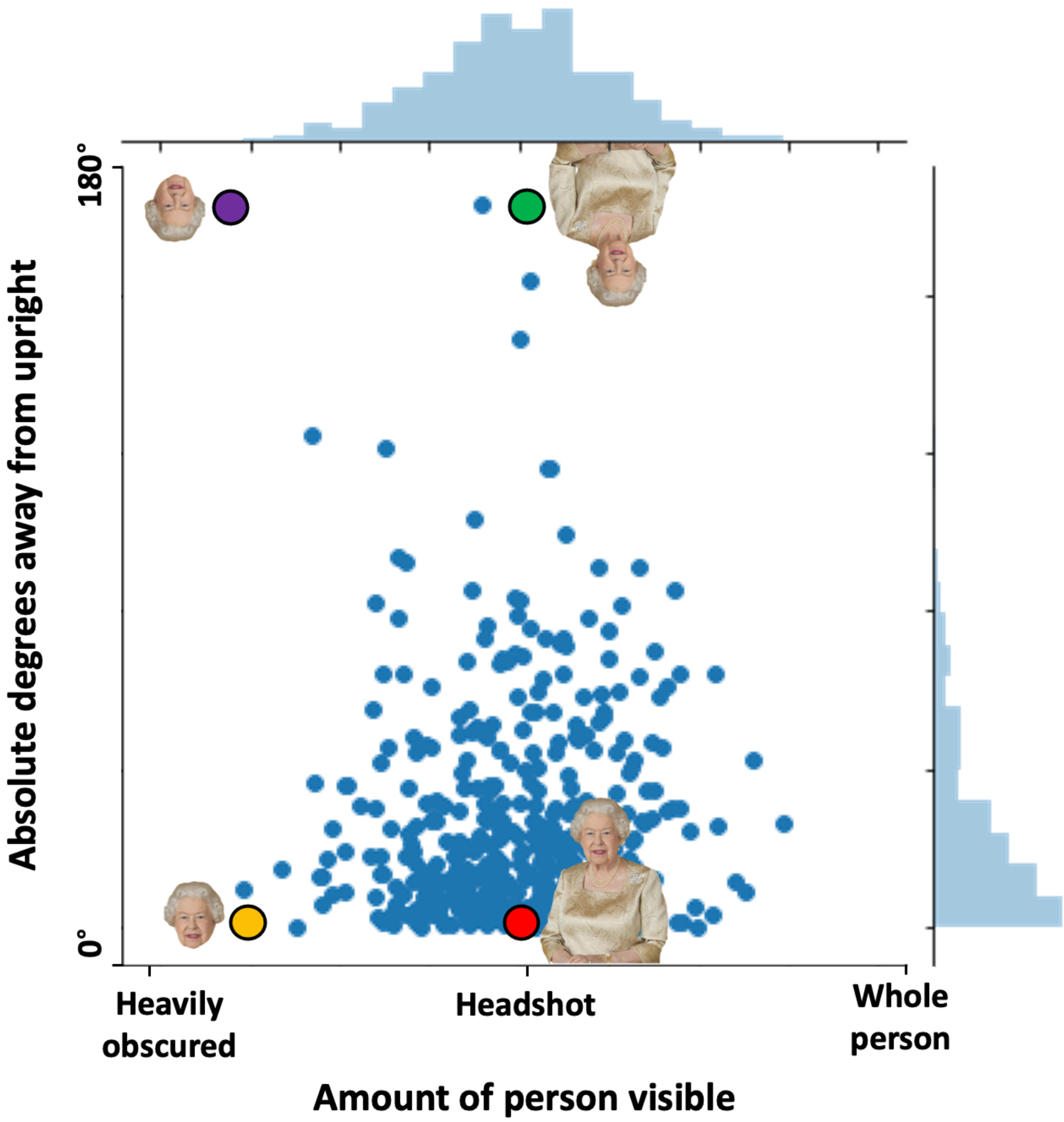



**Figure 4.9.** Possible locations within face space where images from different trials types were encoded. Proximity to previously encoded instances determines likelihood of accurate recognition.

I stated above that familiar and unfamiliar faces are processed qualitatively differently (Johnston & Edmonds, 2009). That idea was born out of the results of chapter 2 which used a matching task. However, in Experiments 1 and 2 of chapter 3, a recognition task rather than a matching task was used. Therefore, in this type of task, the process by which familiar and unfamiliar faces were processed was not qualitatively different, as both were encoded into face space and attempted to be matched to an existing visual person representation. As with the famous faces, the upright unfamiliar faces were encoded into face space and because they were unfamiliar, they were not matched to a visual person representation and could be accurately rejected as unfamiliar, regardless of whether the image was whole or cropped. Inversion causes the visual system to become more error-prone (Valentine, 1991) and therefore when the unfamiliar faces were inverted, again whether cropped or whole, participants were slightly less accurate. Although familiar and unfamiliar faces underwent the same kind of processing in Experiments 1 and 2 and were therefore not processed in qualitatively different ways, the presence (familiar faces) or absence (unfamiliar faces) of visual person representations does reflect a qualitative difference in underlying representations.

Experiments 5 and 6 utilised an old-new memory task. According to the present proposed idea of the spectrum of familiarity, as the unfamiliar faces in Experiment 6 were only seen once in the exposure phase, they would be classed as unfamiliar, and they would be processed qualitatively differently from the famous faces, as participants had previously seen many instances of the famous identities and formed robust visual person representations prior to the experiment. In the test phase, for famous whole upright, whole inverted and cropped upright faces, participants were able to recognise them and therefore accurately say they had seen them before.

However, as the inverted cropped famous images had only the internal features (most of the face and body removed) and were inverted, they did not resemble the previous instances in the visual person representation, were encoded further away and were therefore much less accurately recognised. Therefore participants were less accurate in recalling if they had seen them.

For unfamiliar faces however, participants viewed the single whole or cropped image in the exposure phase and began to form a new visual person representation. This would have been weak and not representative, and of a qualitatively different nature to a familiar visual person representation. In the test phase, both the whole and cropped upright unfamiliar faces had equivalently weak representations prone to forgetting, producing similar decreases in recognition and therefore memory recall. The inverted images, both for cropped and whole images, would have been encoded further away in face space, and as the representations were so small and weak, they could not generalise to match the image producing decreases in accuracy. This is in line with Bruce (1982) and Hancock et al. (2000) who find that unfamiliar face recognition is highly image bound.

The results of chapters 2 and 3 combine to further our understanding of face familiarity. Both experiments lend support to the idea that faces are learnt by observing within-person variation and cohering this together into a visual person (face) representation. Chapter 2 suggests that the formation of the representation is easier than previously thought, as high variation images were still bound together without the aid of top-cues, producing significantly higher matching accuracy compared to novel identities (Experiments 2 and 4). Furthermore, the chapter suggests that faces are learnt with ease and speed, producing a qualitative difference in processing after forming a face representation, as faces where even a small amount of variation had been seen (e.g. just 10 stills from a video) were subsequently matched significantly more accurately than novel identities (Experiments 1-4). Chapter 3 suggests that face representations may be more helpfully thought of as visual person representations, also containing visual information about the external facial features and body as well as the face. It suggests that the face inversion effect in the literature may be stimulus bound, a result of harshly cropped inverted images not resembling the previous observed instances stored in the visual person representation, causing large decreases in accuracy. Inversion of unfamiliar faces may cause medium (Experiments 5 and 6) or small (Experiments 1 and 2) inversion effects depending on whether the task involves processing unfamiliar faces in relation to familiar faces and so producing higher accuracy (Experiments 1 and 2) or not (Experiments 5 and 6). While face familiarity was not the original focus of this thesis, it is clear that unfamiliar and familiar faces are qualitatively different categories and this affects all areas of face perception, including research on how faces are learnt and modulating whether the face inversion effect is present or not.

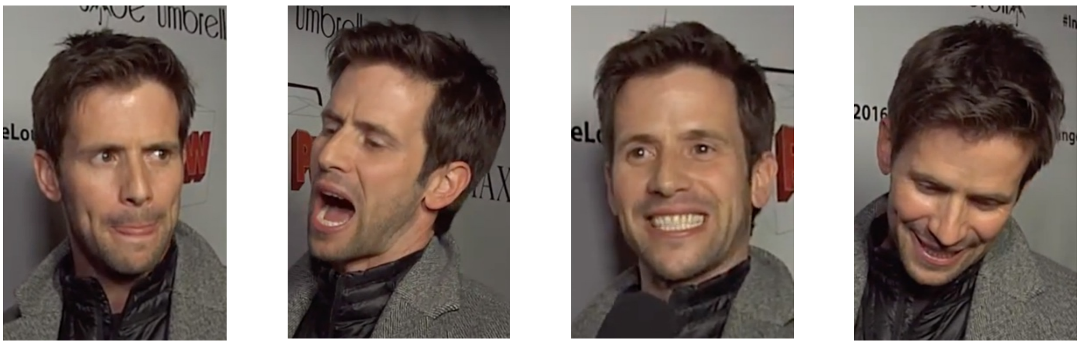
# Appendices

### Appendix A: Further stimuli examples from Chapter 2, Experiments 1 and 2

Low variation images:

High variation images:

Low variation images:



High variation images:

## Appendix B: Further stimuli examples from Chapter 2, Experiments 3 and 4

Low variation images:

High variation images:

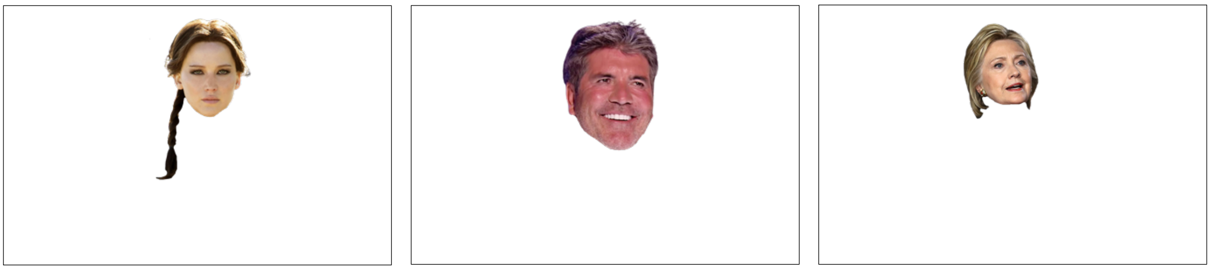
Low variation images:

High variation images:

## Appendix C: Further stimuli examples from Chapter 3, Experiment 1

Familiar Identities Whole Images

Familiar Identities Cropped Images



Unfamiliar Identities Whole Images

Unfamiliar Identities Cropped Images

Familiar Buildings Whole Images

Familiar Buildings Cropped Images

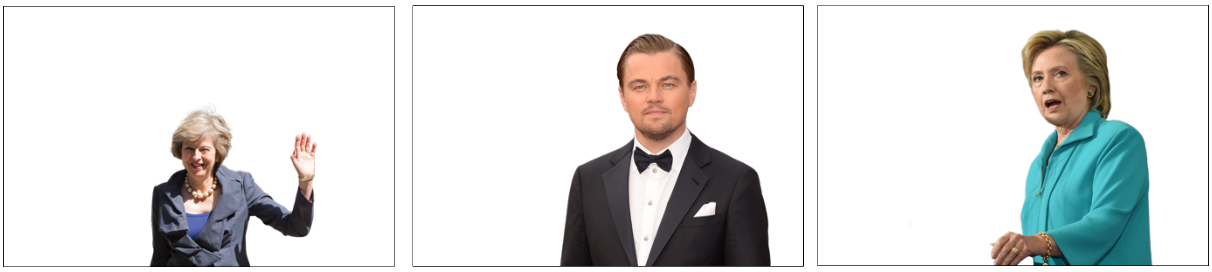
Unfamiliar Buildings Whole Images

Unfamiliar Buildings Cropped Images

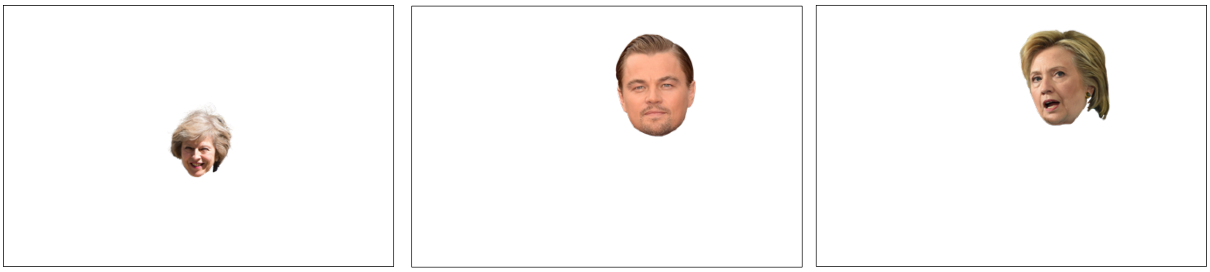
**Appendix D: Further stimuli examples from Chapter 3, Experiment 2**

Familiar Identities Whole Images


Familiar Identities Whole Bodies



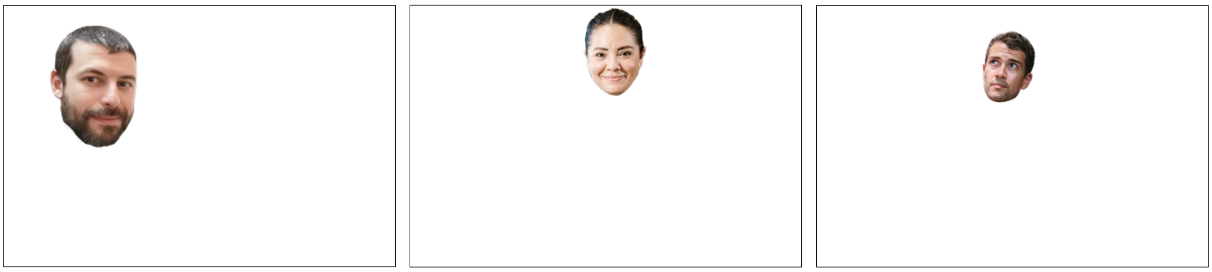
Familiar Identities Cropped Images



Unfamiliar Identities Whole Images

Unfamiliar Identities Whole Bodies

Unfamiliar Identities Cropped Images



## Appendix E: Further stimuli examples from Chapter 3, Experiment 3 and 4

Familiar Identities Whole Images


Familiar Identities Cropped Images

## Appendix F: Further stimuli examples from Chapter 3, Experiment 5

Unfamiliar Identities Whole Images

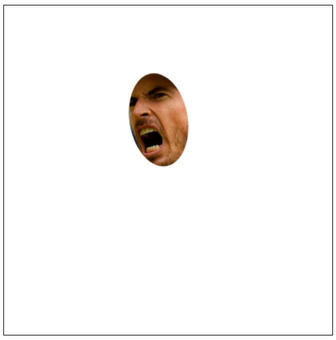
Unfamiliar Identities Cropped Images

## Appendix G: Further stimuli examples from Chapter 3, Experiment 6

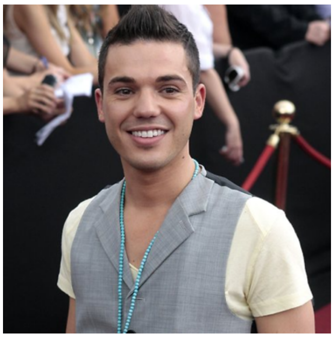
Familiar Identities Whole Images

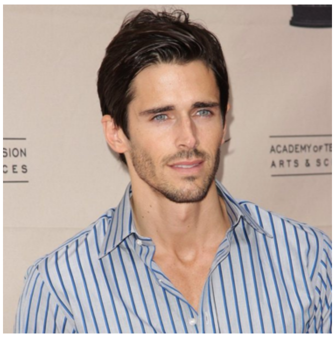


Familiar Identities Cropped Images

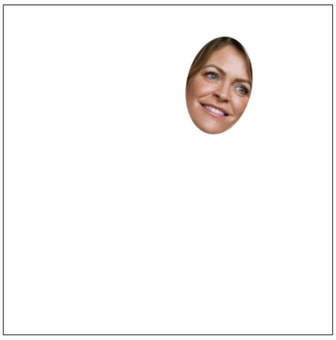




Unfamiliar Identities Whole Images





Unfamiliar Identities Cropped Images


# References


Ambrus, G. G., Eick, C. M., Kaiser, D., & Kovács, G. (2021). Getting to Know You: Emerging Neural Representations during Face Familiarization. *The Journal of Neuroscience*, *41*(26), 5687–5698. https://doi.org/10.1523/JNEUROSCI.2466-20.2021

Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology*, *68*(10), 2041–2050. https://doi.org/10.1080/17470218.2014.1003949

Andrews, T. J., & Schluppeck, D. (2004). Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. *NeuroImage*, *21*(1), 91–98. https://doi.org/10.1016/j.neuroimage.2003.08.023

Ashworth, A. R. S., Vuong, Q. C., Rossion, B., & Tarr, M. J. (2008). Recognizing rotated faces and Greebles: What properties drive the face inversion effect? *Visual Cognition*, *16*(6), 754–784. https://doi.org/10.1080/13506280701381741

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575–589. https://doi.org/10.1016/S0022-5371(75)80045-4

Balas, B. J., Nelson, C. A., Westerlund, A., Vogel-Farley, V., Riggins, T., & Kuefner, D. (2010). Personal familiarity influences the processing of upright and inverted faces in infants. *Frontiers in Human Neuroscience*, *4*(1), 1–6. https://doi.org/10.3389/neuro.09.001.2010

Bartlett, J. C., & Searcy, J. (1993). Inversion and Configuration of Faces. *Cognitive Psychology*, *25*(3), 281–316. https://doi.org/10.1006/cogp.1993.1007

Barton, J. J. S., Radcliffe, N., Cherkasova, M. V., Edelman, J., & Intriligator, J. M. (2006). Information Processing during Face Recognition: The Effects of Familiarity, Inversion, and Morphing on Scanning Fixations. *Perception*, *35*(8), 1089–1105. https://doi.org/10.1068/p5547

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565. https://doi.org/10.1162/jocn.1996.8.6.551

Bernstein, M., Oron, J., Sadeh, B., & Yovel, G. (2014). An Integrated Face–Body Representation in the Fusiform Gyrus but Not the Lateral Occipital Cortex. *Journal of Cognitive Neuroscience*, *26*(11), 2469–2478. https://doi.org/10.1162/jocn_a_00639

Bindemann, M., & Sandford, A. (2011). Me, Myself, and I: Different Recognition Rates for Three Photo-IDs of the Same Person. *Perception*, *40*(5), 625–627. https://doi.org/10.1068/p7008

Bodamer, J. (1947). Die Prosop-Agnosie: Die Agnosie des Physiognomieerkennens. *Archiv für Psychiatrie und Nervenkrankheiten Vereinigt mit Zeitschrift für die Gesamte Neurologie und Psychiatrie*, *179*, 6–53. https://doi.org/10.1007/BF00352849

Bombari, D., Mast, F. W., & Lobmaier, J. S. (2009). Featural, Configural, and Holistic Face-Processing Strategies Evoke Different Scan Patterns. *Perception*, *38*(10), 1508–1521. https://doi.org/10.1068/p6117

Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, *10*(5), 527–536. https://doi.org/10.1080/13506280244000168

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-Racial Identification. *Personality and Social Psychology Bulletin*, *15*(1), 19–25. https://doi.org/10.1177/0146167289151002

Brandman, T., & Yovel, G. (2012). A face inversion effect without a face. *Cognition*, *125*(3), 365–372. https://doi.org/10.1016/j.cognition.2012.08.001

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*(1), 105–116. https://doi.org/10.1111/j.2044-8295.1982.tb01795.x

Bruce, V. (1994). Stability from Variation: The Case of Face Recognition the M.D. Vernon Memorial Lecture. *The Quarterly Journal of Experimental Psychology Section A*, *47*(1), 5–28. https://doi.org/10.1080/14640749408401141

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, A. (1993). Sex Discrimination: How Do We Tell the Difference between Male and Female Faces? *Perception*, *22*(2), 131–152. https://doi.org/10.1068/p220131

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*(4), 339–360. https://doi.org/10.1037/1076-898X.5.4.339

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, *7*(3), 207–218. https://doi.org/10.1037/1076-898X.7.3.207

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

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

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, *51*(3), 256–284. https://doi.org/10.1016/j.cogpsych.2005.06.003

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity From Variation: Representations of Faces Derived From Multiple Instances. *Cognitive Science*, *40*(1), 202–223. https://doi.org/10.1111/cogs.12231

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments Against a Configural Processing Account of Familiar Face Recognition. *Perspectives on Psychological Science*, *10*(4), 482–496. https://doi.org/10.1177/1745691615583129

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. https://doi.org/10.1111/1467-9280.00144

Busigny, T., Mayer, E., & Rossion, B. (2013). Prosopagnosia. In *The behavioral and cognitive neurology of stroke* (pp. 231–244). Cambridge University Press.

Butcher, N., & Lander, K. (2013). The differential role of motion in recognition of familiar and unfamiliar faces. *The Proceedings of the Annual Convention of the Japanese Psychological Association*, *77*(0), 1AM-098-1AM – 098. https://doi.org/10.4992/pacjpa.77.0_1AM-098

Campbell, A., & Tanaka, J. W. (2018). Inversion Impairs Expert Budgerigar Identity Recognition: A Face-Like Effect for a Nonface Object of Expertise. *Perception*, *47*(6), 647–659. https://doi.org/10.1177/0301006618771806

Civile, C., McLaren, R., & McLaren, I. P. L. (2016). The Face Inversion Effect: Roles of First and Second-Order Configural Information. *The American Journal of Psychology*, *129*(1), 23–35. https://doi.org/10.5406/amerjpsyc.129.1.0023

Clutterbuck, R., & Johnston, R. A. (2002). Exploring Levels of Face Familiarity by Using an Indirect Face-Matching Measure. *Perception*, *31*(8), 985–994. https://doi.org/10.1068/p3335

Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition*, *11*(7), 857–869. https://doi.org/10.1080/13506280444000021

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology*, *17*(1), 97–116. https://doi.org/10.1080/09541440340000439

Collin, C. A., & Byrne, S. (2010). The bilateral field advantage in a sequential face–name matching task with famous and nonfamous faces. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *64*(1), 25–32. https://doi.org/10.1037/a0016312

Collishaw, S. M., & Hole, G. J. (2000). Featural and Configurational Processes in the Recognition of Faces of Different Familiarity. *Perception*, *29*(8), 893–909. https://doi.org/10.1068/p2949

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied*, *14*(2), 139–150. https://doi.org/10.1037/1076-898X.14.2.139

Devue, C., & de Sena, S. (2023). The impact of stability in appearance on the development of facial representations. *Cognition*, *239*(2023), 1–13. https://doi.org/10.1016/j.cognition.2023.105569

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. https://doi.org/10.1037/0096-3445.115.2.107

Dowsett, A. J., Sandford, A., & Burton, A. M. (2016). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *Quarterly Journal of Experimental Psychology*, *69*(1), 1–10. https://doi.org/10.1080/17470218.2015.1017513

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

Dunn, J. D., Kemp, R. I., & White, D. (2018). Search templates that incorporate within-face variation improve visual search for faces. *Cognitive Research: Principles and Implications*, *3*(1), 37. https://doi.org/10.1186/s41235-018-0128-1

Dyer, A. G., Neumeyer, C., & Chittka, L. (2005). Honeybee (Apis mellifera) vision can discriminate between and recognise images of human faces. *Journal of Experimental Biology*, *208*(24), 4709–4714. https://doi.org/10.1242/jeb.01929

Ekman, P., & Oster, H. (1979). Facial Expressions of Emotion. *Annual Review of Psychology*, *30*(1), 527–554. https://doi.org/10.1146/annurev.ps.30.020179.002523

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of Familiar and Unfamiliar Faces from Internal and External Features: Some Implications for Theories of Face Recognition. *Perception*, *8*(4), 431–439. https://doi.org/10.1068/p080431

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 628–634. https://doi.org/10.1037/0096-1523.21.3.628

Farah, M. J., Wilson, K. D., Maxwell Drain, H., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research*, *35*(14), 2089–2093. https://doi.org/10.1016/0042-6989(94)00273-O

Foi, A., & Boracchi, G. (2016). Foveated Nonlocal Self-Similarity. *International Journal of Computer Vision*, *120*(1), 78–110. https://doi.org/10.1007/s11263-016-0898-1

Foster, C., Zhao, M., Bolkart, T., Black, M. J., Bartels, A., & Bülthoff, I. (2021). Separated and overlapping neural coding of face and body identity. *Human Brain Mapping*, *42*(13), 4242–4260. https://doi.org/10.1002/hbm.25544

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. https://doi.org/10.1068/p3012

Gao, C., Conte, S., Richards, J. E., Xie, W., & Hanayik, T. (2019). The neural sources of N170: Understanding timing of activation in face-selective areas. *Psychophysiology*, *56*(6), e13336. https://doi.org/10.1111/psyp.13336

Gauthier, I., & Tarr, M. J. (1997). Becoming a “Greeble” Expert: Exploring Mechanisms for Face Recognition. *Vision Research*, *37*(12), 1673–1682. https://doi.org/10.1016/S0042-6989(96)00286-6

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform ‘face area’ increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*(6), 568–573. https://doi.org/10.1038/9224

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The Fusiform “Face Area” is Part of a Network that Processes Faces at the Individual Level. *Journal of Cognitive Neuroscience*, *12*(3), 495–504. https://doi.org/10.1162/089892900562165

Gerlach, C., Kühn, C. D., Mathiassen, A. B., Kristensen, C. L., & Starrfelt, R. (2023). The face inversion effect or the face upright effect? *Cognition*, *232*, 105335. https://doi.org/10.1016/j.cognition.2022.105335

Ghuman, A. S., McDaniel, J. R., & Martin, A. (2010). Face Adaptation without a Face. *Current Biology*, *20*(1), 32–36. https://doi.org/10.1016/j.cub.2009.10.077

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

Gruppuso, V., Lindsay, D. S., & Masson, M. E. J. (2007). I'd know that face anywhere! *Psychonomic Bulletin & Review*, *14*(6), 1085–1089. https://doi.org/10.3758/BF03193095

Hadad, B.-S., Schwartz, S., & Binur, N. (2019). Reduced perceptual specialization in autism: Evidence from the other-race face effect. *Journal of Experimental Psychology: General*, *148*(3), 588–594. https://doi.org/10.1037/xge0000550

Hahn, C. A., O'Toole, A. J., & Phillips, P. J. (2016). Dissecting the time course of person recognition in natural viewing environments. *British Journal of Psychology*, *107*(1), 117–134. https://doi.org/10.1111/bjop.12125

Han, H., Otto, C., & Jain, A. K. (2013). Age estimation from face images: Human vs. machine performance. *2013 International Conference on Biometrics (ICB)*, 1–8. https://doi.org/doi: 10.1109/ICB.2013.6613022

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330–337. https://doi.org/10.1016/S1364-6613(00)01519-9

Hassaballah, M., & Aly, S. (2015). Face recognition: Challenges, achievements and future directions. *IET Computer Vision*, *9*(4), 614–626. https://doi.org/10.1049/iet-cvi.2014.0084

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

Hayes, S. M., Baena, E., Truong, T.-K., & Cabeza, R. (2010). Neural Mechanisms of Context Effects on Face Recognition: Automatic Binding and Context Shift Decrements. *Journal of Cognitive Neuroscience*, *22*(11), 2541–2554. https://doi.org/10.1162/jocn.2009.21379

Heisz, J. J., & Shedden, J. M. (2009). Semantic Learning Modifies Perceptual Face Processing. *Journal of Cognitive Neuroscience*, *21*(6), 1127–1134. https://doi.org/10.1162/jocn.2009.21104

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, *15*(4), 445–464. https://doi.org/10.1002/acp.718

Hill, H., & Bruce, V. (1996). The effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 986–1004. https://doi.org/10.1037/0096-1523.22.4.986

Hole, G. J. (1994). Configurational Factors in the Perception of Unfamiliar Faces. *Perception*, *23*(1), 65–74. https://doi.org/10.1068/p230065

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of Geometric Distortions on Face-Recognition Performance. *Perception*, *31*(10), 1221–1240. https://doi.org/10.1068/p3252

Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Experimental Brain Research*, *176*(2), 356–373. https://doi.org/10.1007/s00221-006-0621-5

Hunnisett, N., & Favelle, S. (2021). Within-person variability can improve the identification of unfamiliar faces across changes in viewpoint. *Quarterly Journal of Experimental Psychology*, *74*(11), 1873–1887. https://doi.org/10.1177/17470218211009771

Jacob, C., Guéguen, N., Boulbry, G., & Ardiccioni, R. (2010). Waitresses' facial cosmetics and tipping: A field experiment. *International Journal of Hospitality Management*, *29*(1), 188–190. https://doi.org/10.1016/j.ijhm.2009.04.003

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

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

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1–2), 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577–596. https://doi.org/10.1080/09658210902976969

Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). `Putting the Face to the Voice'. *Current Biology*, *13*(19), 1709–1714. https://doi.org/10.1016/j.cub.2003.09.005

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kawai, N. (2011). Attentional shift by eye gaze requires joint attention: Eye gaze cues are unique to shift attention. *Japanese Psychological Research*, *53*(3), 292–301. https://doi.org/10.1111/j.1468-5884.2011.00470.x

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the Displacement of Facial Features in Negative and Inverted Images. *Perception*, *19*(4), 531–543. https://doi.org/10.1068/p190531

Kovács, G. (2020). Getting to Know Someone: Familiarity, Person Recognition, and Identification in the Human Brain. *Journal of Cognitive Neuroscience*, *32*(12), 2205–2225. https://doi.org/10.1162/jocn_a_01627

Kramer, R. S. S., Jenkins, R., Young, A. W., & Burton, A. M. (2017). Natural variability is essential to learning new faces. *Visual Cognition*, *25*(4–6), 470–476. https://doi.org/10.1080/13506285.2016.1242522

Kramer, R. S. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition*, *172*, 46–58. https://doi.org/10.1016/j.cognition.2017.12.005

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, *104*(51), 20600–20605. https://doi.org/10.1073/pnas.0705654104

Lander, K., & Bruce, V. (2003). The role of motion in learning new faces. *Visual Cognition*, *10*(8), 897–912. https://doi.org/10.1080/13506280344000149

Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology*, *15*(1), 101–116. https://doi.org/10.1002/1099-0720(200101/02)15:1<101::AID-ACP697>3.0.CO;2-7

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory & Cognition*, *27*(6), 974–985. https://doi.org/10.3758/BF03201228

Lander, K., & Chuang, L. (2005). Why are moving faces easier to recognize? *Visual Cognition*, *12*(3), 429–442. https://doi.org/10.1080/13506280444000382

Lander, K., & Davies, R. (2007). Exploring the role of characteristic motion when learning new faces. *Quarterly Journal of Experimental Psychology*, *60*(4), 519–526. https://doi.org/10.1080/17470210601117559

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in Holistic Face Processing Following Early Visual Deprivation. *Psychological Science*, *15*(11), 762–768. https://doi.org/10.1111/j.0956-7976.2004.00753.x

Lee, K. J., & Perrett, D. (1997). Presentation-Time Measures of the Effects of Manipulations in Colour Space on Discrimination of Famous Faces. *Perception*, *26*(6), 733–752. https://doi.org/10.1068/p260733

Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., & Rao, S. M. (2000). Neural Systems Underlying the Recognition of Familiar and Newly Learned Faces. *The Journal of Neuroscience*, *20*(2), 878–886. https://doi.org/10.1523/JNEUROSCI.20-02-00878.2000

Liu, C. H., Bhuiyan, Md. A.-A., Ward, J., & Sui, J. (2009). Transfer between pose and illumination training in face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 939–947. https://doi.org/10.1037/a0013710

Liu, J., Li, J., Feng, L., Li, L., Tian, J., & Lee, K. (2014). Seeing Jesus in toast: Neural and behavioral correlates of face pareidolia. *Cortex*, *53*, 60–77. https://doi.org/10.1016/j.cortex.2014.01.013

Liu, Y., Stowe, M. H., Bello, D., Sparer, J., Gore, R. J., Cullen, M. R., Redlich, C. A., & Woskie, S. R. (2008). Skin Exposure to Aliphatic Polyisocyanates in the Auto Body Repair and Refinishing Industry: III. A Personal Exposure Algorithm. *The Annals of Occupational Hygiene*, *53*(1), 33–40. https://doi.org/10.1093/annhyg/men070

Loftus, E. F. (1980). Impact of expert psychological testimony on the unreliability of eyewitness identification. *Journal of Applied Psychology*, *65*(1), 9–15. https://doi.org/10.1037/0021-9010.65.1.9

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about 'weapon focus.' *Law and Human Behavior*, *11*(1), 55–62. https://doi.org/10.1007/BF01044839

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 77–100. https://doi.org/10.1037/0096-1523.34.1.77

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271. https://doi.org/10.1037/0033-295X.87.3.252

Marzi, T., & Viggiano, M. P. (2007). Interplay between familiarity and orientation in face processing: An ERP study. *International Journal of Psychophysiology*, *65*(3), 182–192. https://doi.org/10.1016/j.ijpsycho.2007.04.003

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260. https://doi.org/10.1016/S1364-6613(02)01903-4

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

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. https://doi.org/10.3758/PBR.16.5.778

McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: A Face-Specific Disorder. *The Quarterly Journal of Experimental Psychology Section A*, *46*(1), 1–10. https://doi.org/10.1080/14640749308401064

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*(4), 865–876. https://doi.org/10.3758/BF03193433

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception & Psychophysics*, *69*(7), 1175–1184. https://doi.org/10.3758/BF03193954

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology*, *94*(3), 339–354. https://doi.org/10.1348/000712603767876262

Menon, N., Kemp, R. I., & White, D. (2018). More than a sum of parts: Robust face recognition by integrating variation. *Royal Society Open Science*, *5*(5), 172381. https://doi.org/10.1098/rsos.172381

Mileva, M., & Burton, A. M. (2018). Smiles in face matching: Idiosyncratic information revealed through a smile improves unfamiliar face matching performance. *British Journal of Psychology*, *109*(4), 799–811. https://doi.org/10.1111/bjop.12318

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural Face Processing Develops more Slowly than Featural Face Processing. *Perception*, *31*(5), 553–566. https://doi.org/10.1068/p3339

Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., & Kleiner-Gathercoal, K. A. (1999). Face Perception During Early Infancy. *Psychological Science*, *10*(5), 419–422. https://doi.org/10.1111/1467-9280.00179

Monti, C., Sozzi, M., Bossi, F., Corbo, M., & Rivolta, D. (2019). Atypical holistic processing of facial identity and expression in a case of acquired prosopagnosia. *Cognitive Neuropsychology*, *36*(7–8), 358–382. https://doi.org/10.1080/02643294.2020.1718071

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 577–581. https://doi.org/10.1037/xhp0000049

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, *6*(6), 261–266. https://doi.org/10.1016/S1364-6613(02)01908-3

Pascalis, O., De Haan, M., & Nelson, C. A. (2002). Is Face Processing Species-Specific During the First Year of Life? *Science*, *296*(5571), 1321–1323. https://doi.org/10.1126/science.1070223

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the Human Body in the Fusiform Gyrus. *Journal of Neurophysiology*, *93*(1), 603–608. https://doi.org/10.1152/jn.00513.2004

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 23–30. https://doi.org/10.1098/rstb.1992.0003

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple Dissociation of Faces, Bodies, and Objects in Extrastriate Cortex. *Current Biology*, *19*(4), 319–324. https://doi.org/10.1016/j.cub.2009.01.007

Pitcher, D., Ianni, G., & Ungerleider, L. G. (2019). A functional dissociation of face-, body- and scene-selective brain areas based on their response to moving and static stimuli. *Scientific Reports*, *9*(1), 8242. https://doi.org/10.1038/s41598-019-44663-9

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS Evidence for the Involvement of the Right Occipital Face Area in Early Face Processing. *Current Biology*, *17*(18), 1568–1573. https://doi.org/10.1016/j.cub.2007.07.063

Psalta, L., Young, A. W., Thompson, P., & Andrews, T. J. (2014). Orientation-sensitivity to facial features explains the Thatcher illusion. *Journal of Vision*, *14*(12), 9–9. https://doi.org/10.1167/14.12.9

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential Sensitivity of Human Visual Cortex to Faces, Letterstrings, and Textures: A Functional Magnetic Resonance Imaging Study. *The Journal of Neuroscience*, *16*(16), 5205–5215. https://doi.org/10.1523/JNEUROSCI.16-16-05205.1996

Rakover, S. S., Bar-On, R. A., & Gliklich, A. (2022). Comparing Inverted Faces to Upright Faces Using Similarity or Mental Rotation. *The American Journal of Psychology*, *135*(2), 161–167. https://doi.org/10.5406/19398298.135.2.03

Rezlescu, C., Chapman, A. F., Susilo, T., & Caramazza, A. (2016). *Large inversion effects are not specific to faces and do not vary with object expertise* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/xzbe5

Rice, A., Phillips, P. J., Natu, V., An, X., & O'Toole, A. J. (2013). Unaware Person Recognition From the Body When Face Identification Fails. *Psychological Science*, *24*(11), 2235–2243. https://doi.org/10.1177/0956797613492986

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, *70*(5), 897–905. https://doi.org/10.1080/17470218.2015.1136656

Ritchie, K. L., Kramer, R. S. S., & Burton, A. M. (2018). What makes a face photo a 'good likeness'? *Cognition*, *170*, 1–8. https://doi.org/10.1016/j.cognition.2017.09.001

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*(1), 34–79. https://doi.org/10.1016/j.cognition.2006.02.008

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, *128*(2), 274–289. https://doi.org/10.1016/j.actpsy.2008.02.003

Russell, R., & Sinha, P. (2007). Real-World Face Recognition: The Importance of Surface Reflectance Properties. *Perception*, *36*(9), 1368–1374. https://doi.org/10.1068/p5779

Schwartz, L., & Yovel, G. (2016). The roles of perceptual and conceptual information in face recognition. *Journal of Experimental Psychology: General*, *145*(11), 1493–1511. https://doi.org/10.1037/xge0000220

Schwartz, L., & Yovel, G. (2019). Learning faces as concepts rather than percepts improves face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1733–1747. https://doi.org/10.1037/xlm0000673

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate Face and Body Selectivity on the Fusiform Gyrus. *The Journal of Neuroscience*, *25*(47), 11055–11059. https://doi.org/10.1523/JNEUROSCI.2621-05.2005

Sekiguchi, T. (2011). Individual differences in face memory and eye fixation patterns during face learning. *Acta Psychologica*, *137*(1), 1–9. https://doi.org/10.1016/j.actpsy.2011.01.014

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion Leads to Quantitative, Not Qualitative, Changes in Face Processing. *Current Biology*, *14*(5), 391–396. https://doi.org/10.1016/j.cub.2004.02.028

Shackelford, T. K., & Larsen, R. J. (1999). Facial attractiveness and physical health. *Evolution and Human Behavior*, *20*(1), 71–76. https://doi.org/10.1016/S1090-5138(98)00036-1

Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. *Proceedings of the IEEE*, *94*(11), 1948–1962. https://doi.org/10.1109/JPROC.2006.884093

Sliwinska, M. W., & Pitcher, D. (2018). TMS demonstrates that both right and left superior temporal sulci are important for facial expression recognition. *NeuroImage*, *183*, 394–400. https://doi.org/10.1016/j.neuroimage.2018.08.025

Sliwinska, M. W., Searle, L. R., Earl, M., O'Gorman, D., Pollicina, G., Burton, A. M., & Pitcher, D. (2022). Face learning via brief real-world social interactions induces changes in face-selective brain areas and hippocampus. *Perception*, *51*(8), 521–538. https://doi.org/10.1177/03010066221098728

Stewart, E. E. M., Valsecchi, M., & Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision*, *20*(12), 2. https://doi.org/10.1167/jov.20.12.2

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

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. https://doi.org/10.1080/14640749308401045

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, *25*(5), 583–592. https://doi.org/10.3758/BF03211301

Thompson, P. (1980). Margaret Thatcher: A New Illusion. *Perception*, *9*(4), 483–484. https://doi.org/10.1068/p090483

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*(4), 471–491. https://doi.org/10.1111/j.2044-8295.1988.tb02747.x

Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161–204. https://doi.org/10.1080/14640749108400966

Valentine, T., & Bruce, V. (1988). Mental rotation of faces. *Memory & Cognition*, *16*(6), 556–566. https://doi.org/10.3758/BF03197057

Wang, Z., Wu, T., Zhang, W., Deng, W., Li, Y., Zhang, L., Sun, Y.-H. P., & Jin, H. (2023). High familiar faces have both eye recognition and holistic processing advantages. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-023-02792-4

White, D., Burton, A. M., Jenkins, R., & Kemp, R. I. (2014). Redesigning photo-ID to improve unfamiliar face matching performance. *Journal of Experimental Psychology: Applied*, *20*(2), 166–173. https://doi.org/10.1037/xap0000009

White, D., Wayne, T., & Varela, V. P. L. (2022). Partitioning natural face image variability emphasises within-identity over between-identity representation for understanding accurate recognition. *Cognition*, *219*, 104966. https://doi.org/10.1016/j.cognition.2021.104966

Wiese, H., Schipper, M., Popova, T., Burton, A. M., & Young, A. W. (2023). Personal familiarity of faces, animals, objects, and scenes: Distinct perceptual and overlapping conceptual representations. *Cognition*, *241*, 105625. https://doi.org/10.1016/j.cognition.2023.105625

Wiese, H., Tüttenberg, S. C., Burton, M., & Young, A. W. (2019). The Sustained Familiarity Effect: A robust neural correlate of familiar face recognition. *Journal of Vision*, *19*(10), 93. https://doi.org/10.1167/19.10.93

Wixted, J. T., & Ebbesen, E. B. (1991). On the Form of Forgetting. *Psychological Science*, *2*(6), 409–415. https://doi.org/10.1111/j.1467-9280.1991.tb00175.x

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. https://doi.org/10.1037/h0027474

Yip, A. W., & Sinha, P. (2002). Contribution of Color to Face Recognition. *Perception*, *31*(8), 995–1003. https://doi.org/10.1068/p3376

Young, A. W., & Burton, A. M. (2017). Recognizing Faces. *Current Directions in Psychological Science*, *26*(3), 212–217. https://doi.org/10.1177/0963721416688114

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational Information in Face Perception. *Perception*, *16*(6), 747–759. https://doi.org/10.1068/p160747

Yovel, G., & O'Toole, A. J. (2016). Recognizing People in Motion. *Trends in Cognitive Sciences*, *20*(5), 383–395. https://doi.org/10.1016/j.tics.2016.02.005